



<https://github.com/wangleiphy/dl4csrc> Slides & Codes

wangleiphy/DL4CSRC: Hitchhiker's guide

Branch: master New pull request Create new file Upload files Find file Clone or download ▾

wangleiphy allow using checkpoint in the flow Latest commit 2920a5d 11 hours ago

1-bp	polish README	14 hours ago
2-isng	allow to select targets	14 hours ago
3-flow	allow using checkpoint in the flow	11 hours ago
assets	upload things by hand	21 hours ago
.gitignore	added flow	3 days ago
README.md	added credit	3 days ago

README.md

Tutorial codes for crash course "Deep Learning for Computational Scientists"
Beijing CSRC, 2018 November 27

- 0. [A hitchhiker's guide](#)
- 1. [Backprop from scratch](#)
- 2. [Differentiable Ising solver](#)
- 3. [Fun with normalizing flows](#)

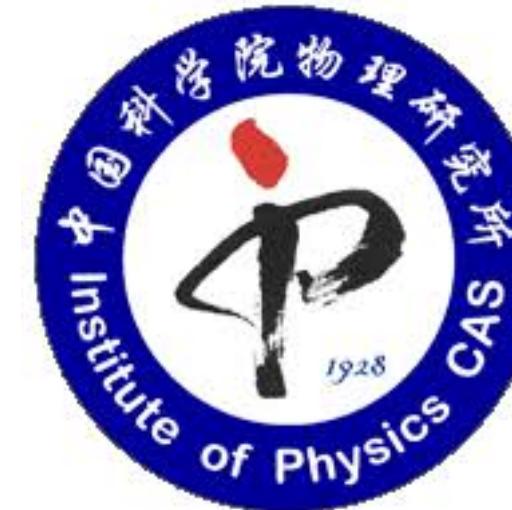
Please send questions and feedbacks to wanglei@iphy.ac.cn. Thank you!

Deep Learning for Computational Scientists

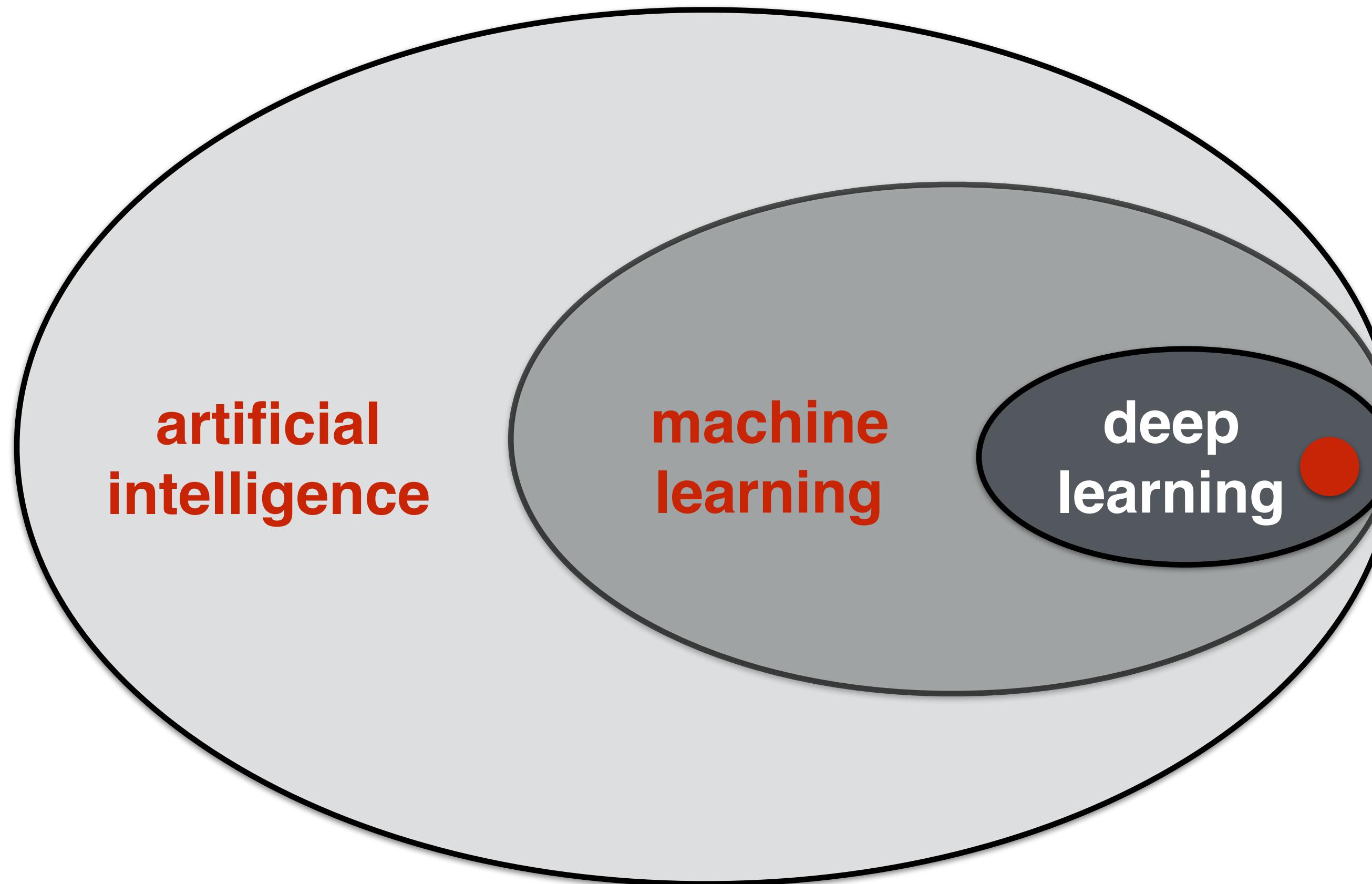
Lei Wang (王磊)

<https://wangleiphy.github.io>

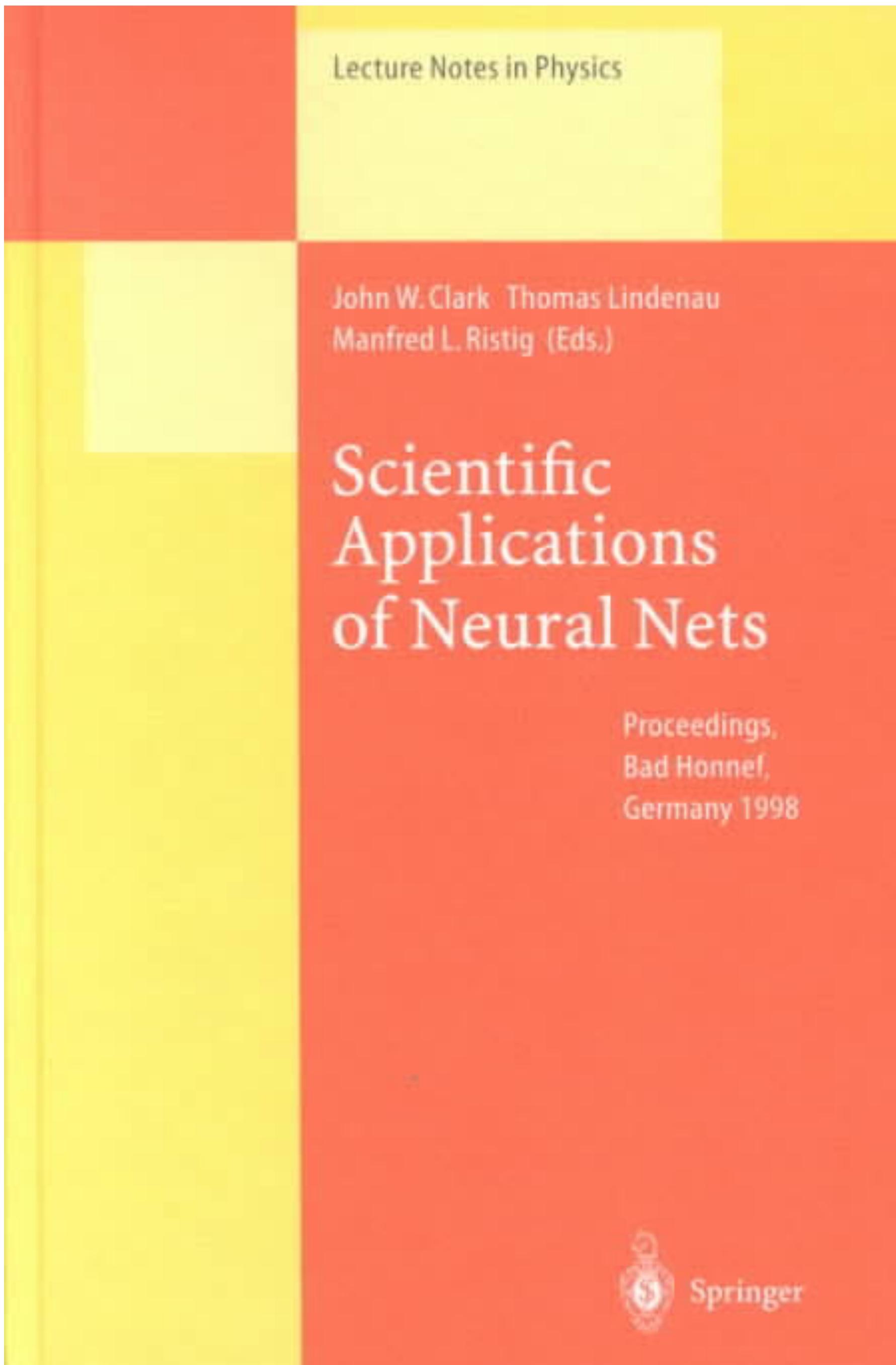
Institute of Physics, Beijing
Chinese Academy of Sciences



Why deep learning ?



**Game changing technology for scientific research
especially computational science**



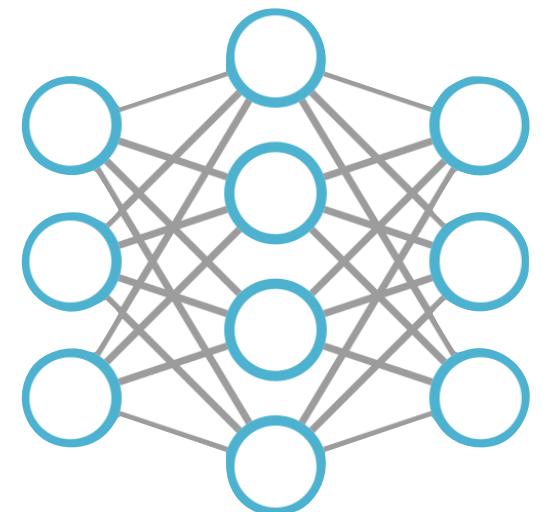
Gem in between this and last hype cycles

8 Doing Science With Neural Nets: Pride and Prejudice

When neural networks re-emerged on the scene in the mid-80s as a new and glamorous computational paradigm, the initial reaction in some sectors of the scientific community was perhaps too enthusiastic and not sufficiently critical. There was a tendency on the part of practitioners to oversell the powers of neural-network or “connectionist” solutions relative to conventional techniques – where conventional techniques can include both traditional theory-rich modeling and established statistical methods. The last five years have seen a correction phase, as some of the practical limitations of neural-network approaches have become apparent, and as scientists have become better acquainted with the wide array of advanced statistical tools that are currently available.

Why now, again ?
[What has changed ?](#)
[What has not ?](#)

Plan



Hitchhiker's guide to deep learning



Secrets behind deep learning

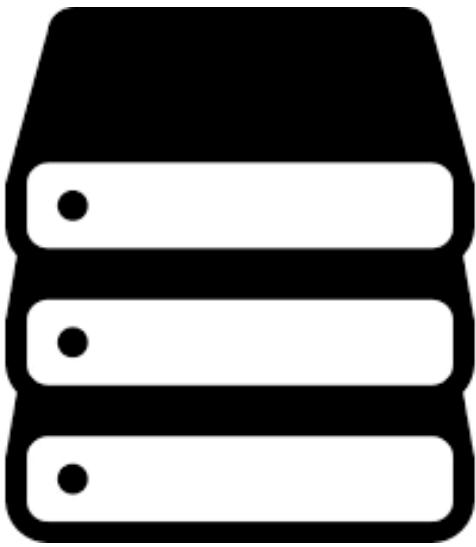


Hands on time

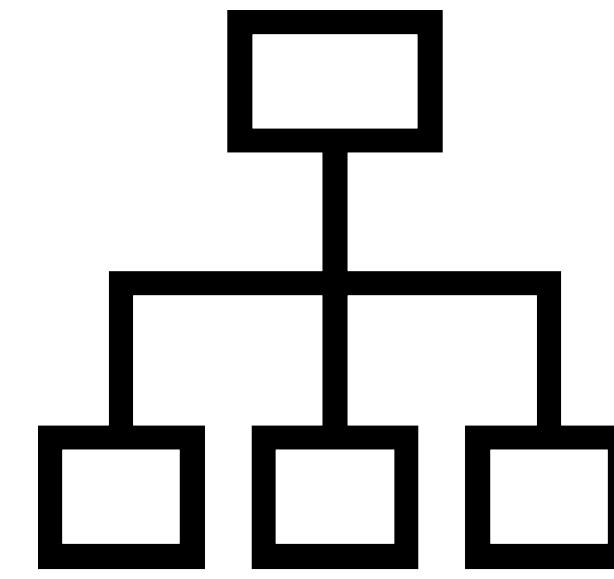
Don't Panic!

Key components

Data



Model



Cost function

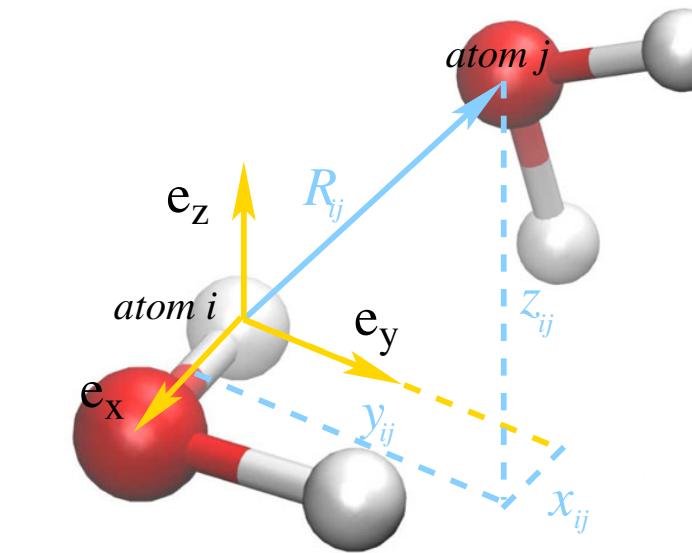
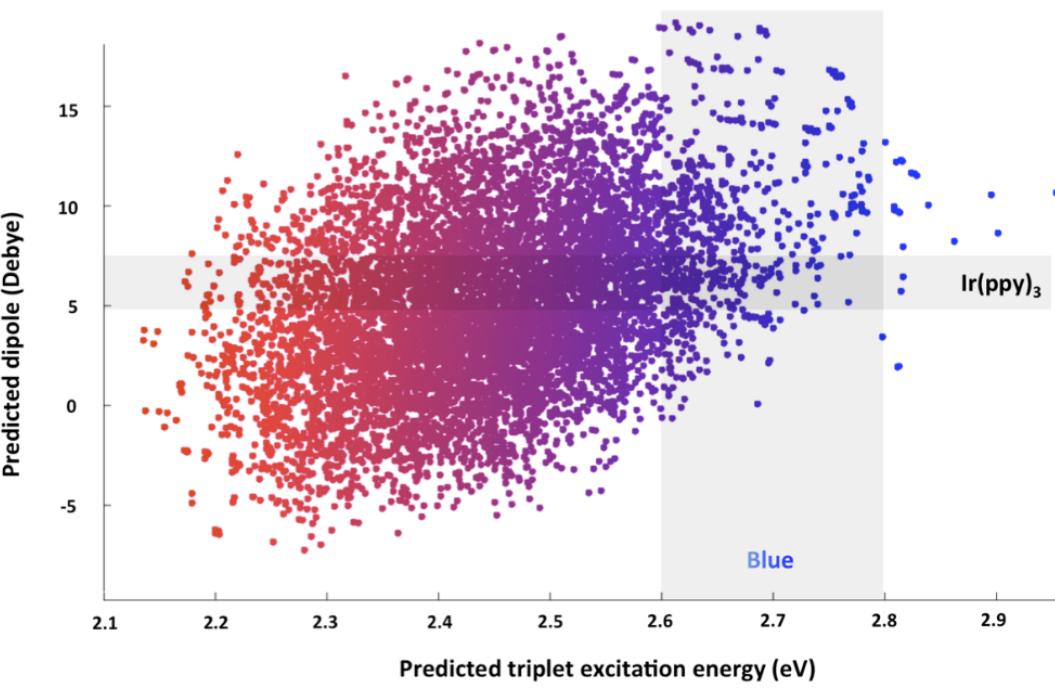


Optimization

$\hat{\theta}$

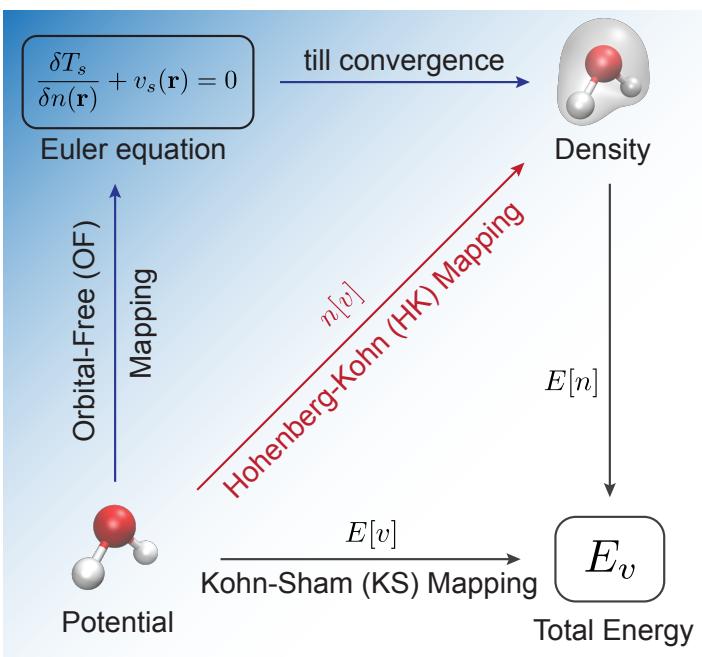
Switch to blackboard

Some applications

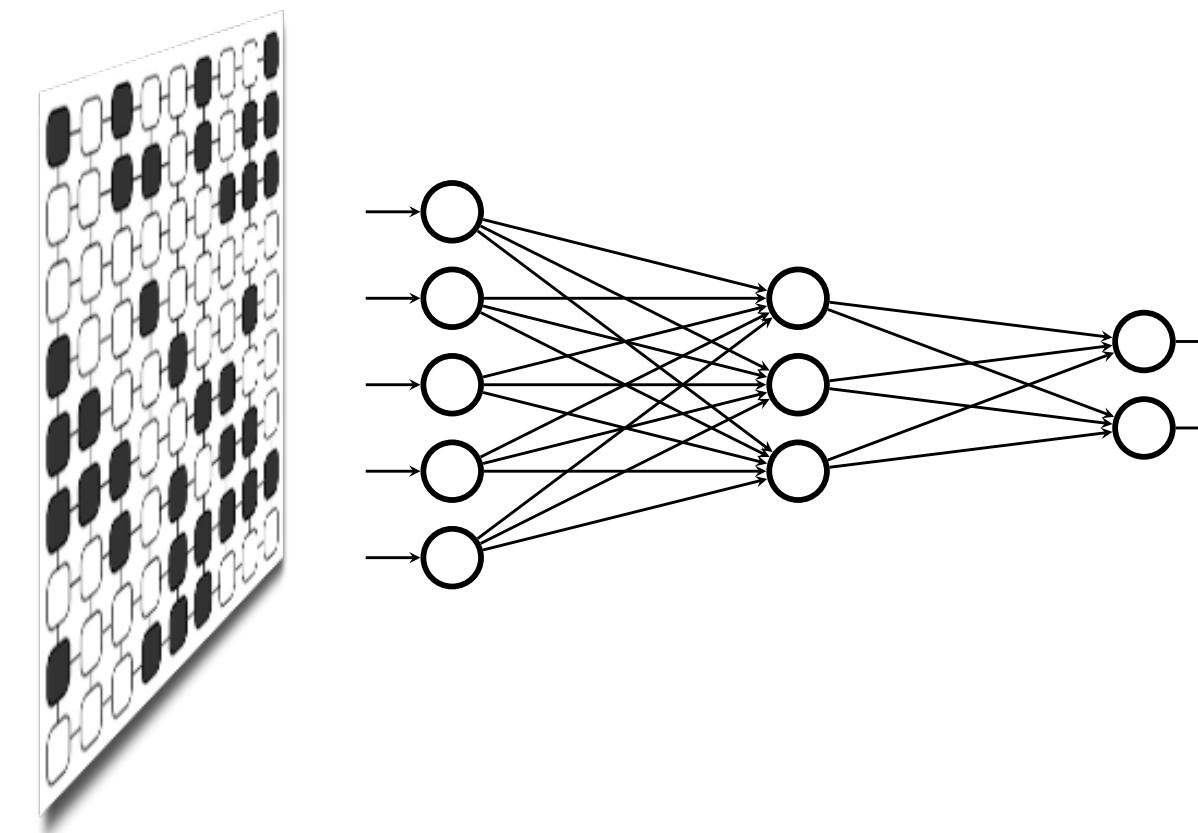


Materials informatics

Molecular simulation

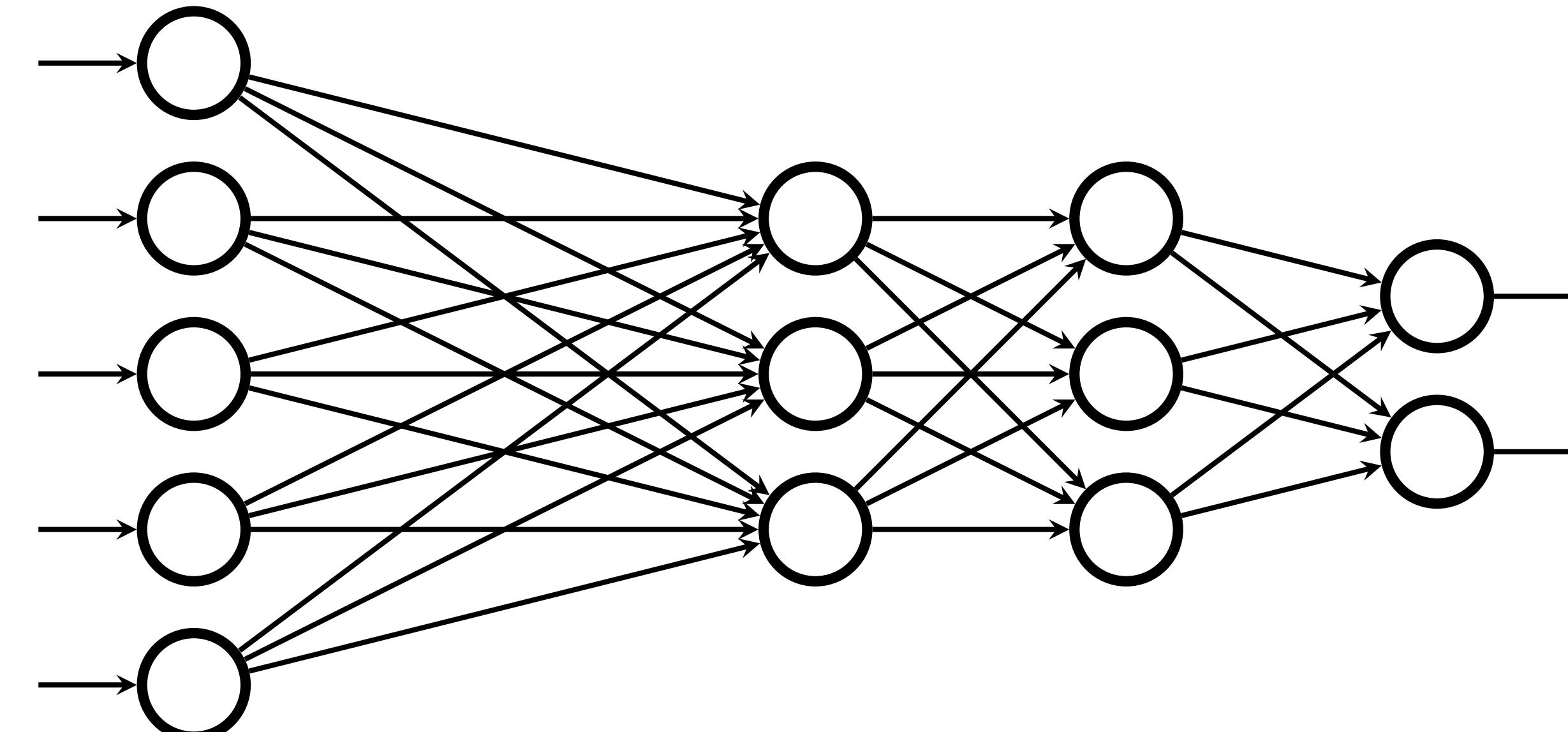
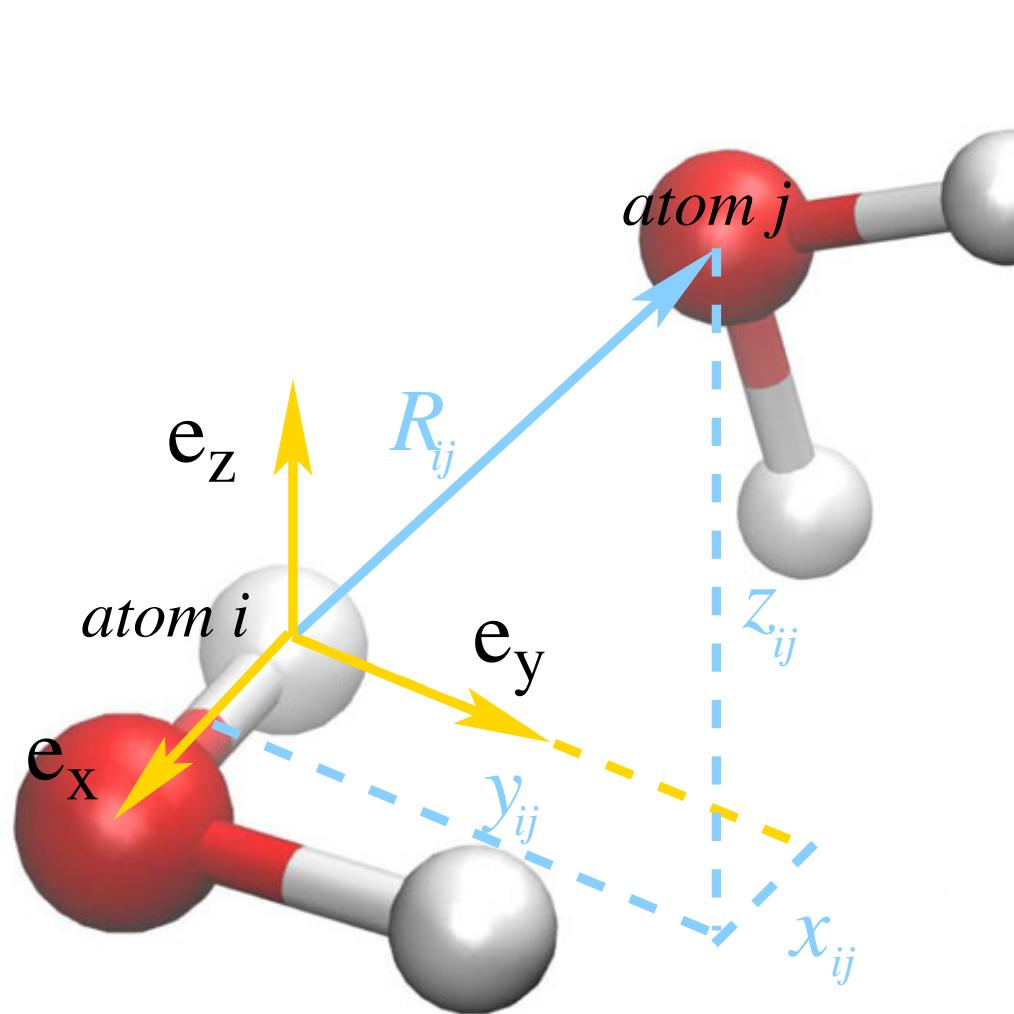


Density functionals

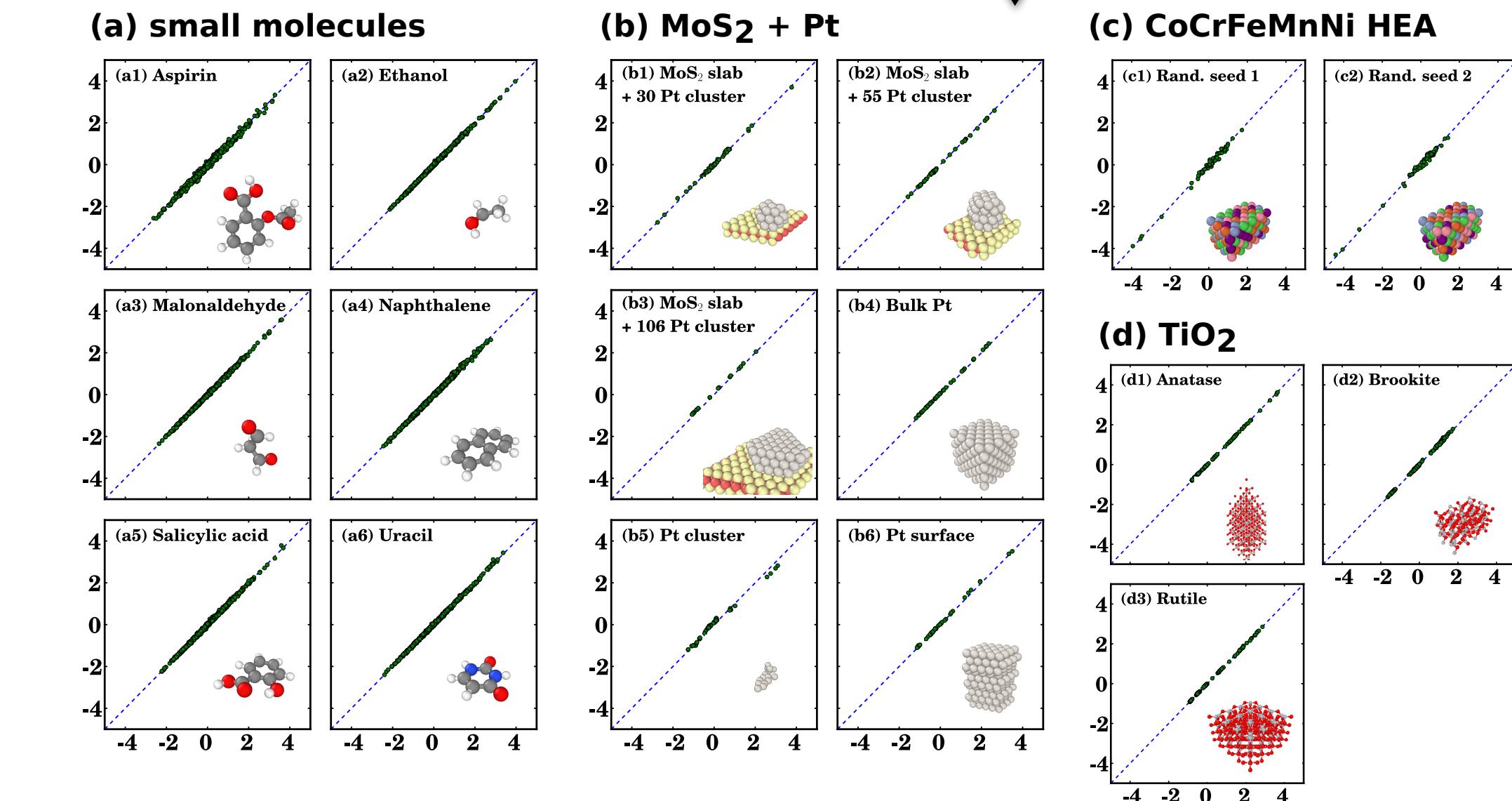


“Phase” recognition

Machine learning energy potential



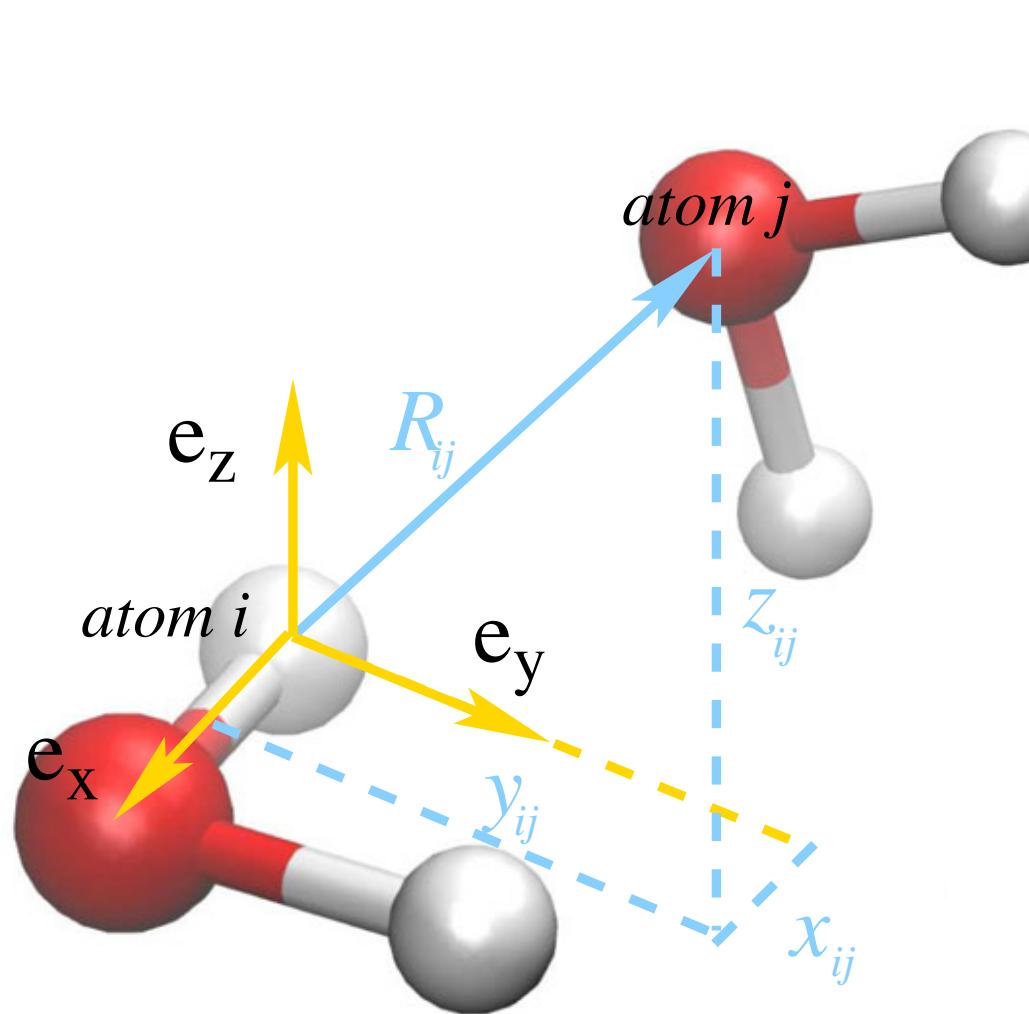
Atom species,
position...



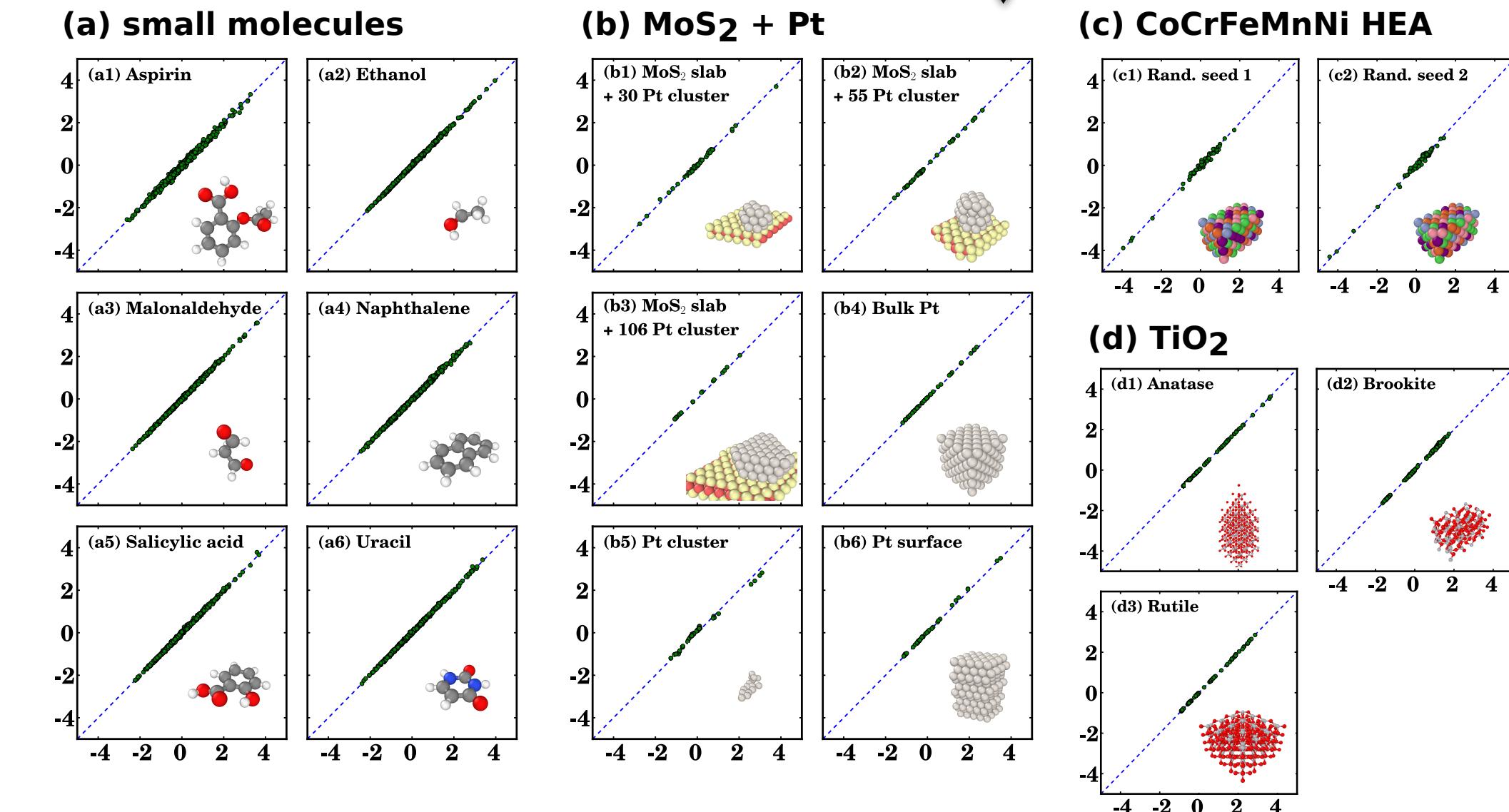
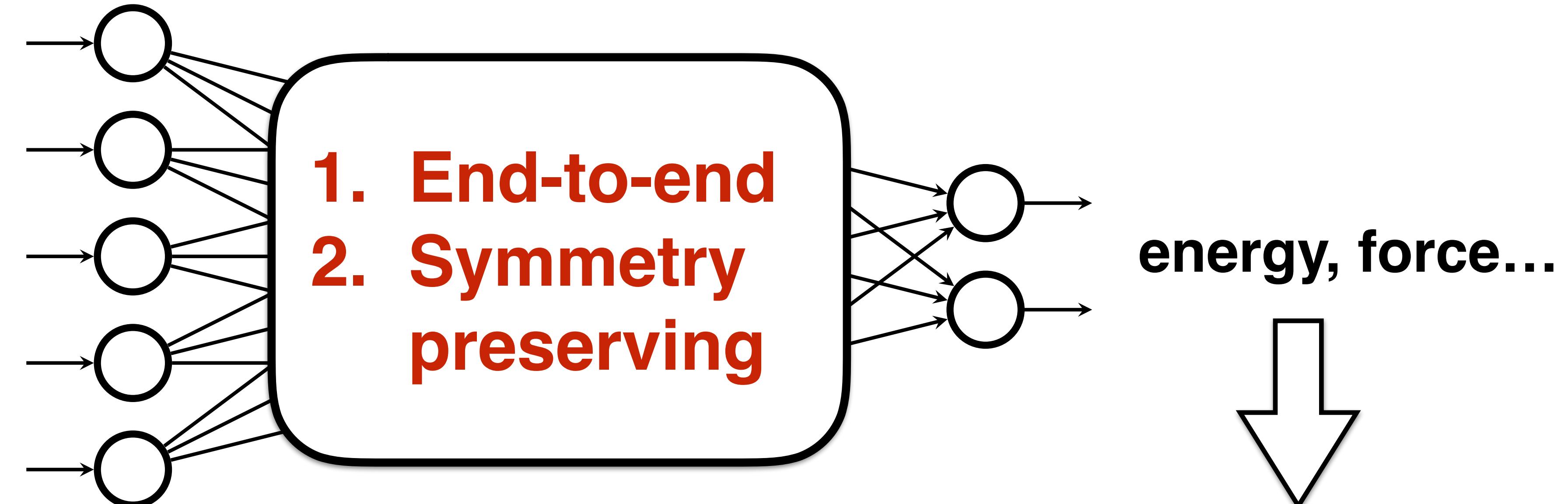
Zhang, Han, Wang, Car, E, PRL 2018

Zhang, Han, Wang, Saidi, Car, E, NIPS 2018

Machine learning energy potential



**Atom species,
position...**

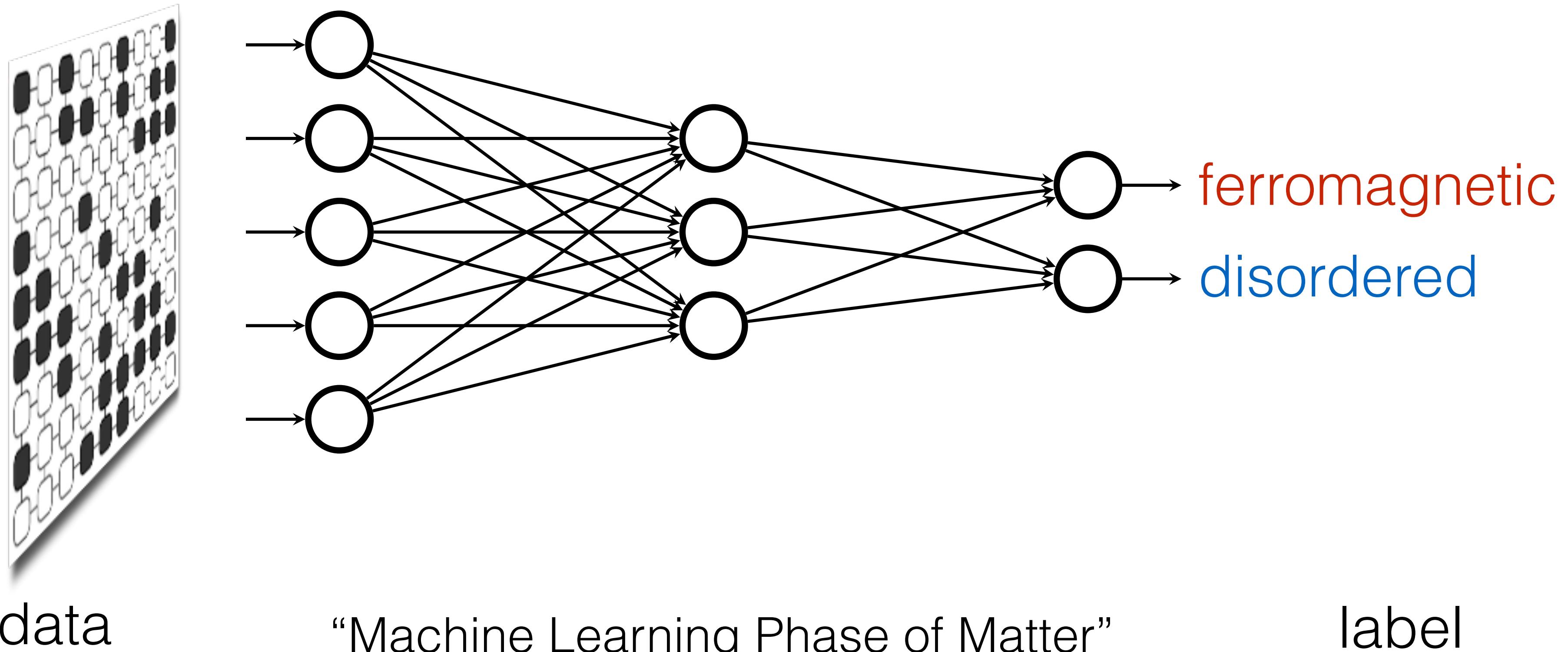


Zhang, Han, Wang, Car, E, PRL 2018

Zhang, Han, Wang, Saidi, Car, E, NIPS 2018

Phase classifications

Ising configurations



Carrasquilla and Melko, 1605.01735

+ many more on quantum spins, fermions, disordered,
topological systems ...

Deep learning is more than function fitting



Discriminative

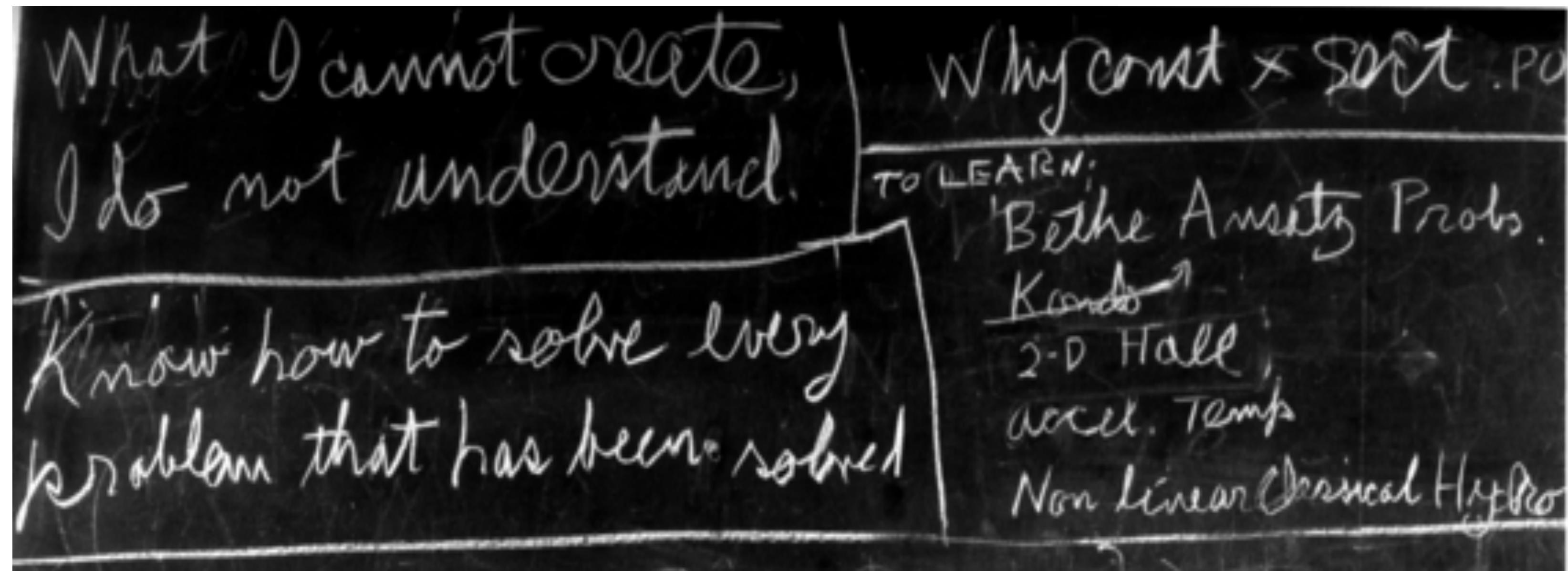
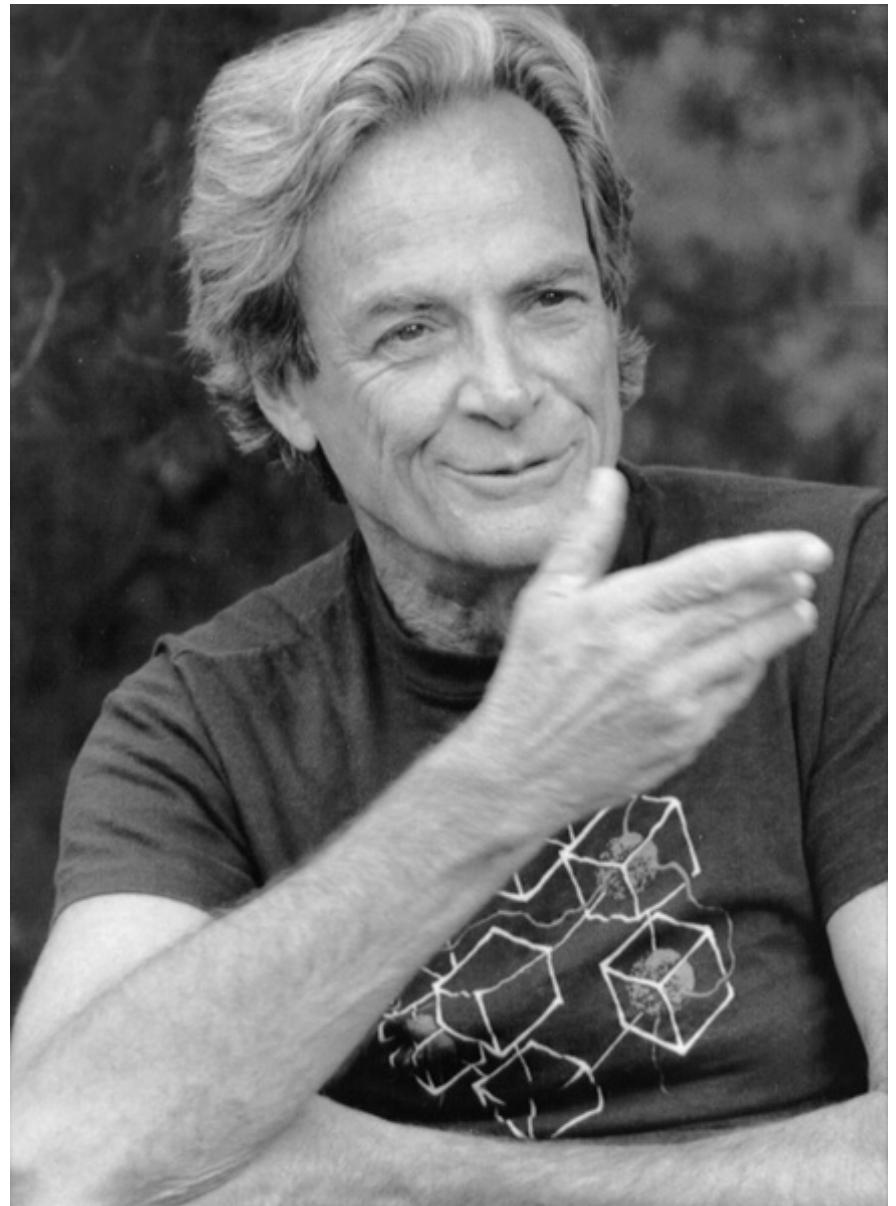
$$y = f(x) \text{ or } p(y | x)$$



Generative

$$p(x, y)$$

Deep learning is more than function fitting



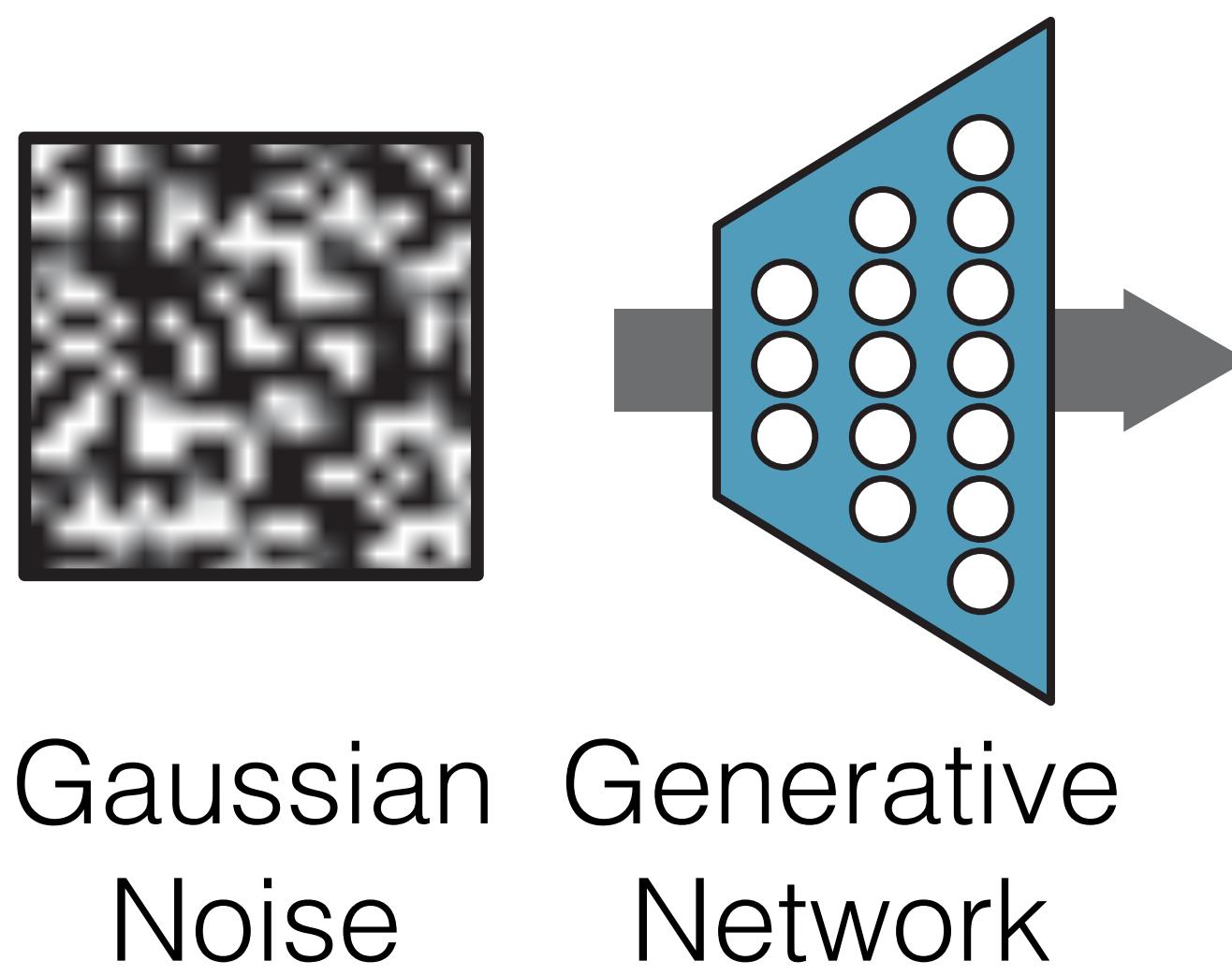
“What I can not create, I do not understand”

Generated Arts

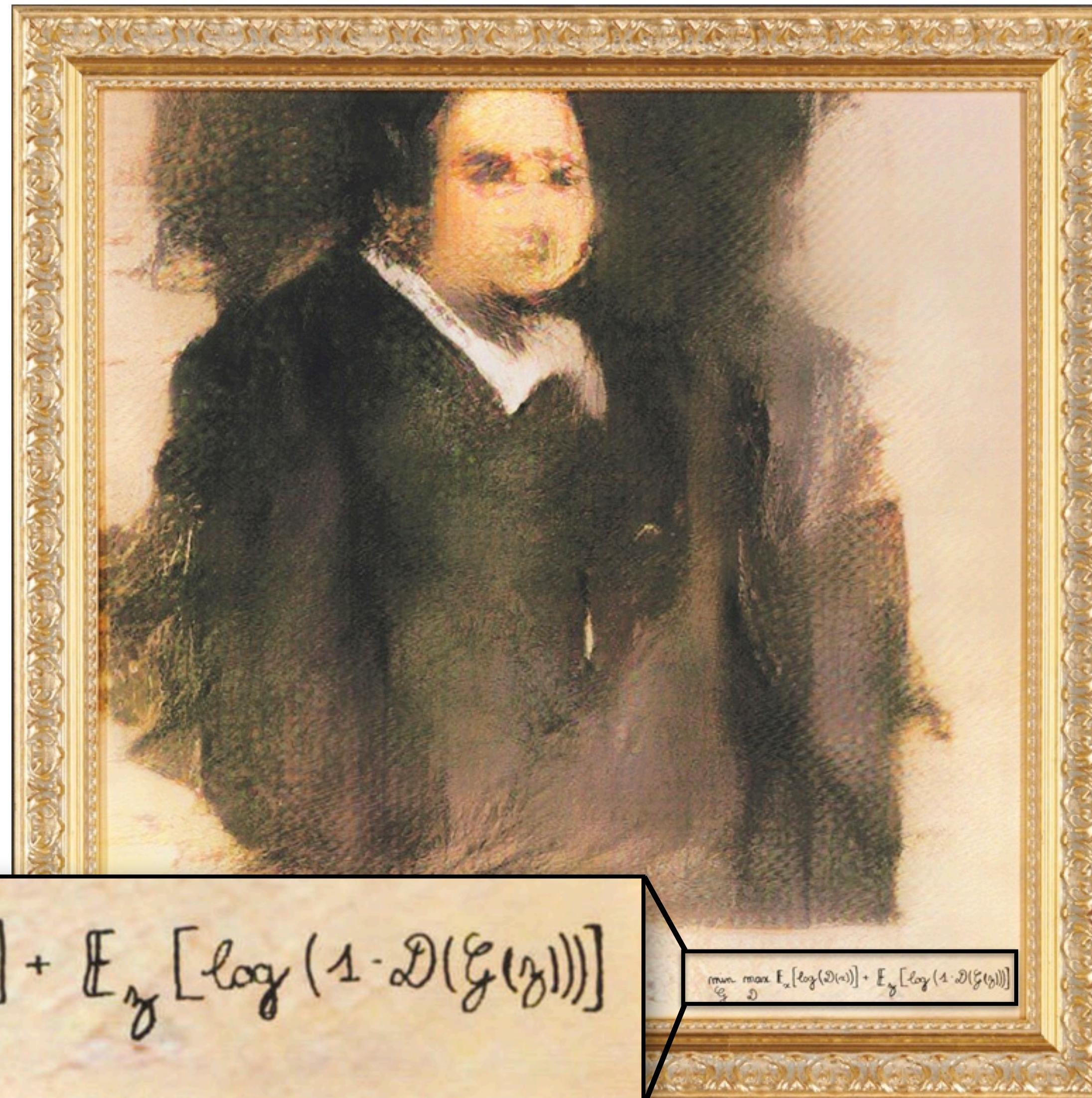


\$432,500
25 October 2018
Christie's New York

Generated Arts



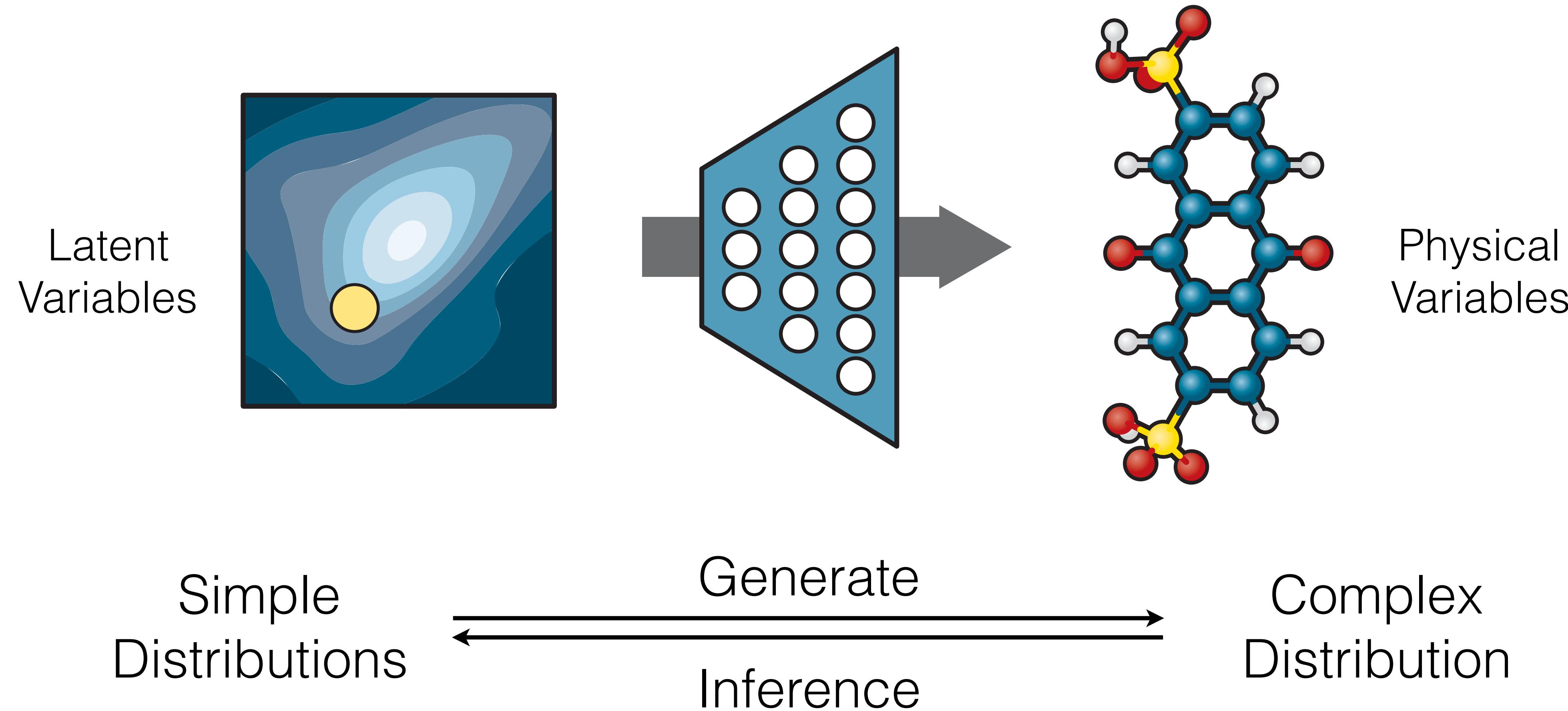
Gaussian Noise Generative Network



$$\min_{\mathcal{G}} \max_{\mathcal{D}} \mathbb{E}_{\mathbf{x}} [\log(\mathcal{D}(\mathbf{x}))] + \mathbb{E}_{\mathbf{z}} [\log(1 - \mathcal{D}(\mathcal{G}(\mathbf{z})))]$$

\$432,500
25 October 2018
Christie's New York

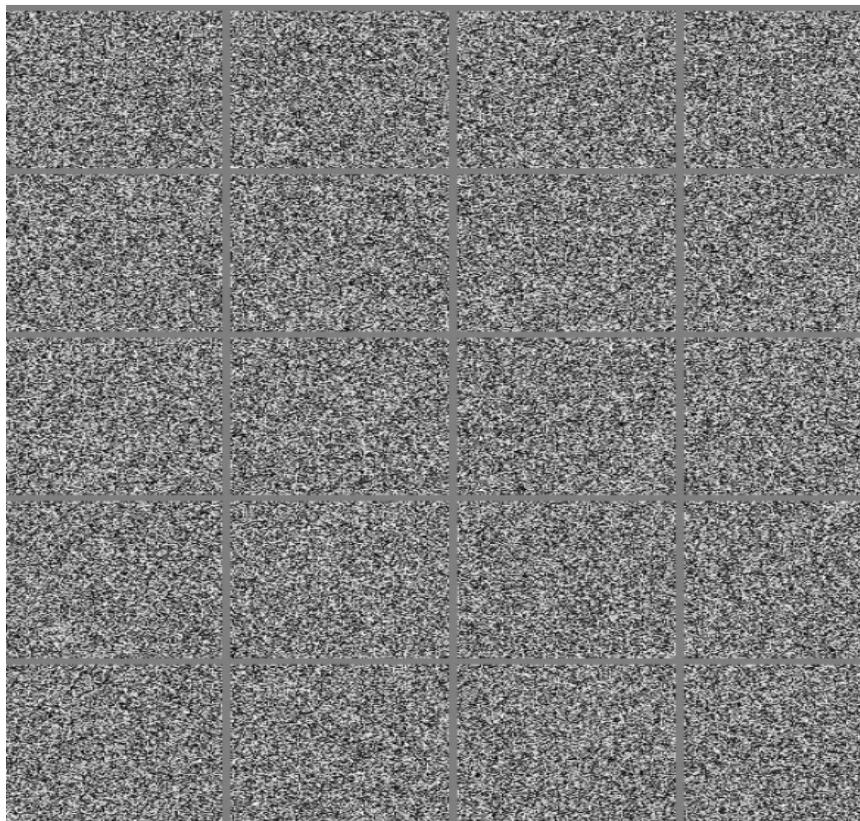
Generate Molecules



Probabilistic Generative Modeling

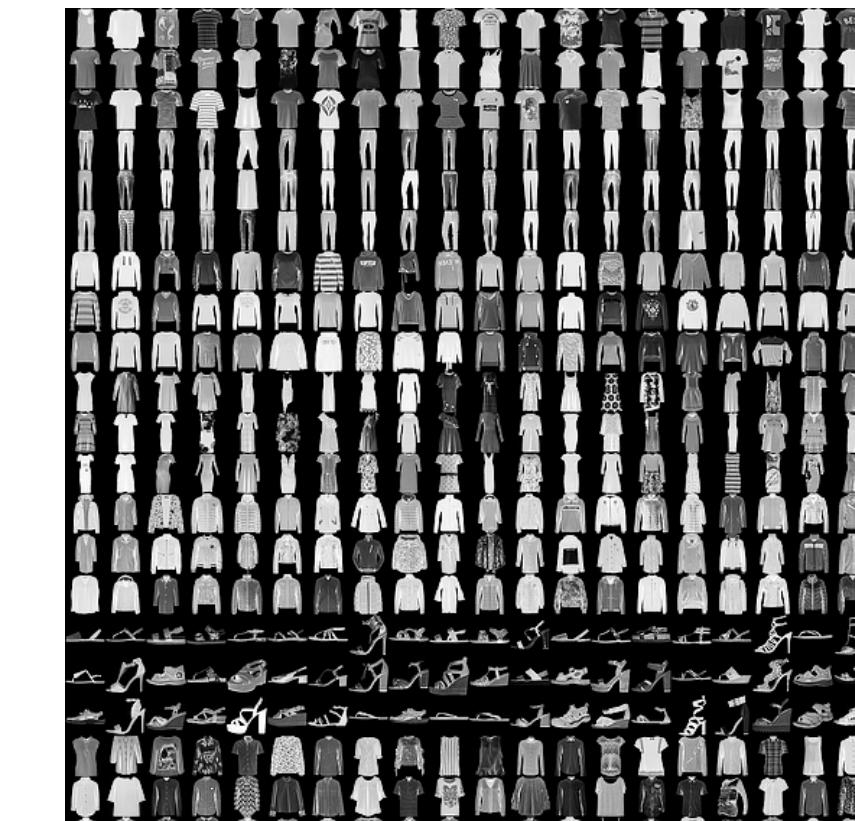
$$p(x)$$

How to express, learn, and sample from a high-dimensional probability distribution ?



“random” images

8	9	0	1	2	3	4	7	8	9	0	1	2	3	4	5	6	7	8	6
4	2	6	4	7	5	5	4	7	8	9	2	9	3	9	3	8	2	0	5
0	1	0	4	2	6	5	3	5	3	8	0	0	3	4	1	5	3	0	8
3	0	6	2	7	1	1	8	1	7	1	3	8	9	7	6	7	4	1	6
7	5	1	7	1	9	8	0	6	9	4	9	9	3	7	1	9	2	2	5
3	7	8	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	0
1	2	3	4	5	6	7	8	9	8	1	0	5	5	1	9	0	4	1	9
3	8	4	7	7	8	5	0	6	5	5	3	3	3	9	8	1	4	0	6
1	0	0	6	2	1	1	3	2	8	8	7	8	4	6	0	2	0	3	6
8	7	1	5	9	9	3	2	4	9	4	4	5	3	2	8	5	9	4	1
6	5	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7
8	9	0	1	2	3	4	5	6	7	8	9	6	4	2	6	4	7	5	5
4	7	8	9	2	9	3	9	3	8	2	0	9	8	0	5	6	0	1	0
4	2	6	5	5	5	4	3	4	1	5	3	0	8	3	0	6	2	7	1
1	8	1	7	1	3	8	5	4	2	0	9	7	6	7	4	1	6	8	4
7	5	1	2	6	7	1	9	8	0	6	9	4	9	9	6	2	3	7	1
9	2	2	5	3	7	8	0	1	2	3	4	5	6	7	8	0	1	2	3
4	5	6	7	8	0	1	2	3	4	5	6	7	8	9	2	1	2	1	3
9	9	8	5	3	7	0	7	7	5	7	9	9	4	7	0	3	4	1	4
4	7	5	8	1	4	8	4	1	8	6	4	6	3	5	7	2	5	9	



“natural” images

Probabilistic modeling

How to
high-d

DEEP LEARNING

Ian Goodfellow, Yoshua Bengio,
and Aaron Courville

from a
solution ?

Page 159

*“... the images encountered in
AI applications occupy a
negligible proportion of
the volume of image space.”*

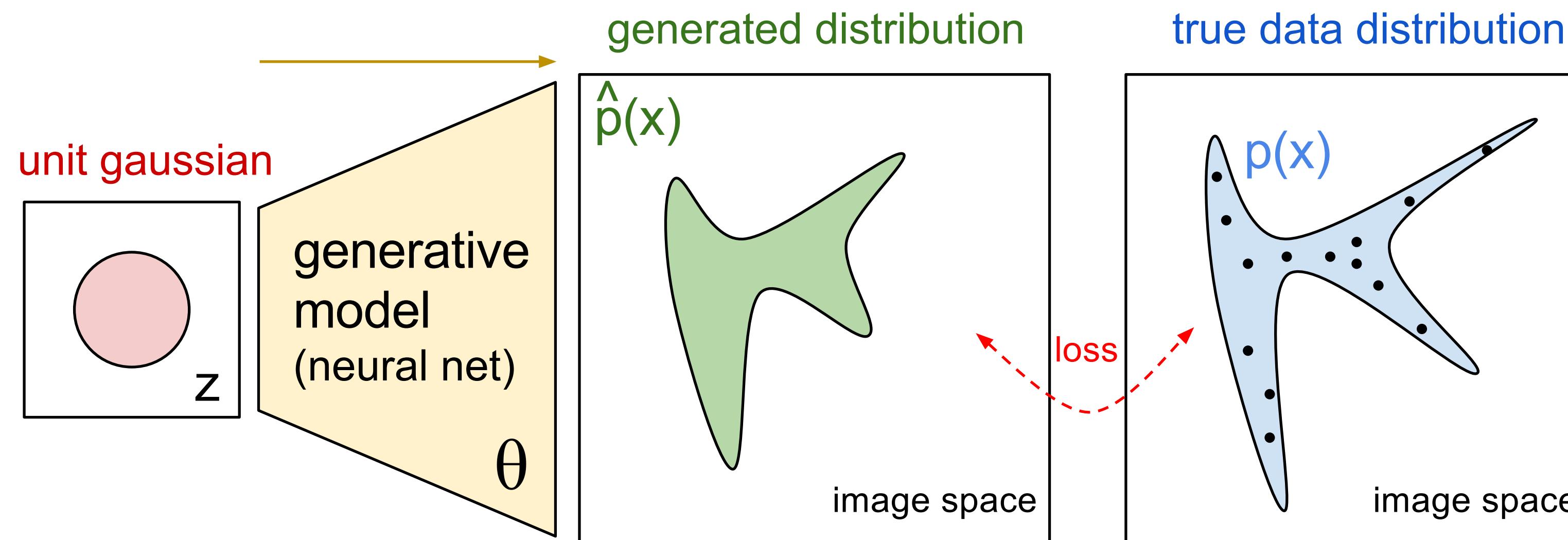
“random”



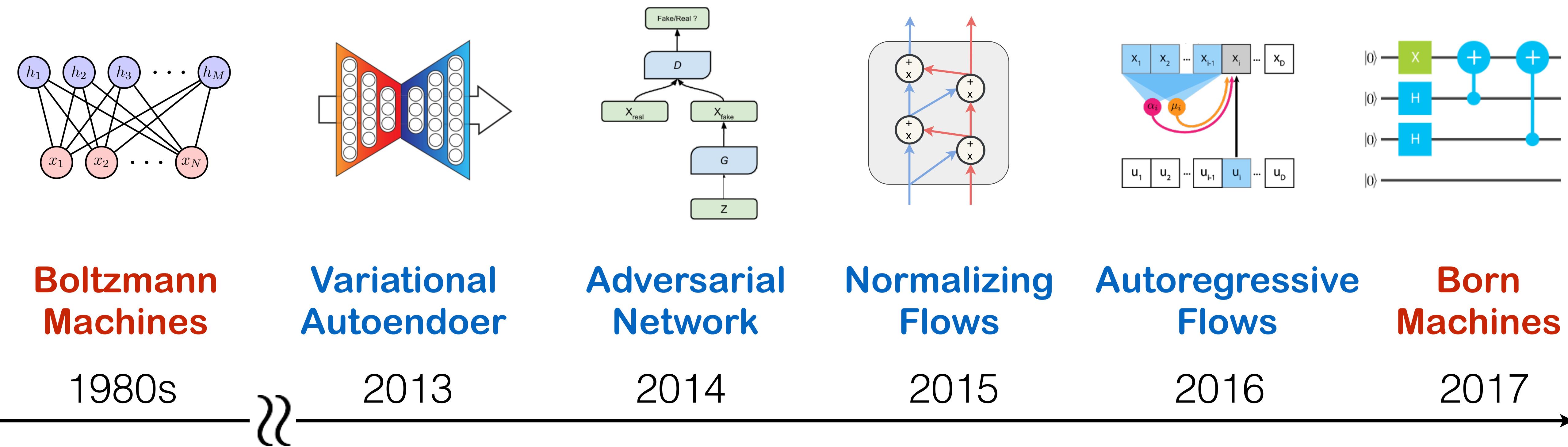
Probabilistic Generative Modeling

$$p(x)$$

How to express, learn, and sample from a high-dimensional probability distribution ?



Timeline of Generative Models



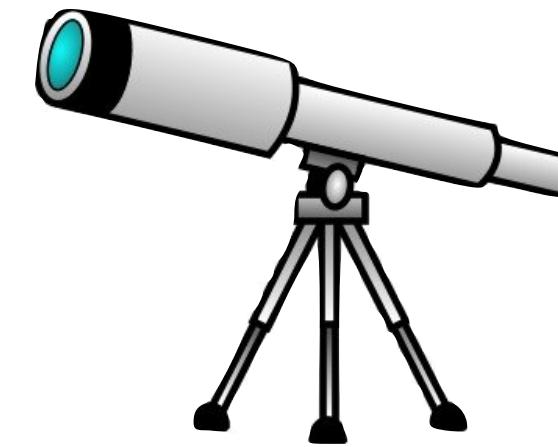
- ① **Leverage the power of modern generative models for physics**
- ② **Statistical, quantum, and fluid mechanics inspired generative models**

Switch to blackboard

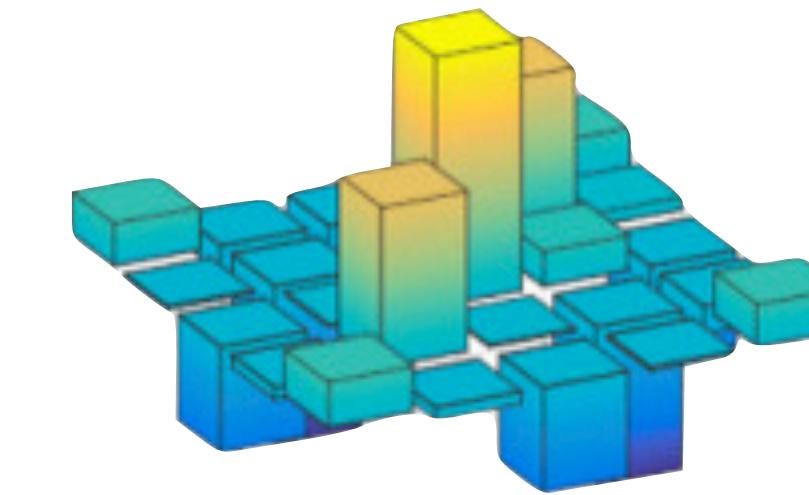
Application of generative models

Ψ

Variational ansatz



Renormalization group

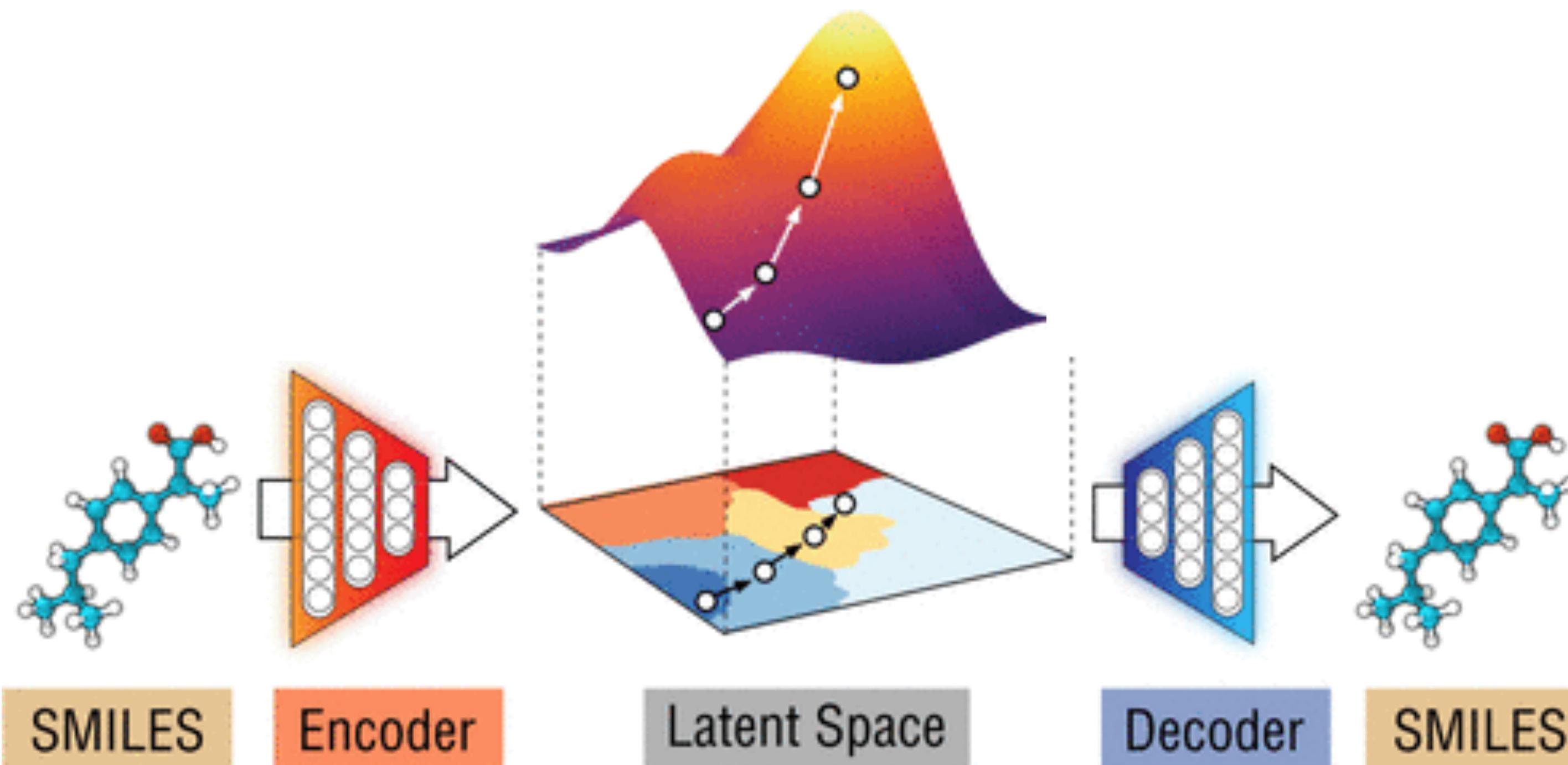


Quantum tomography

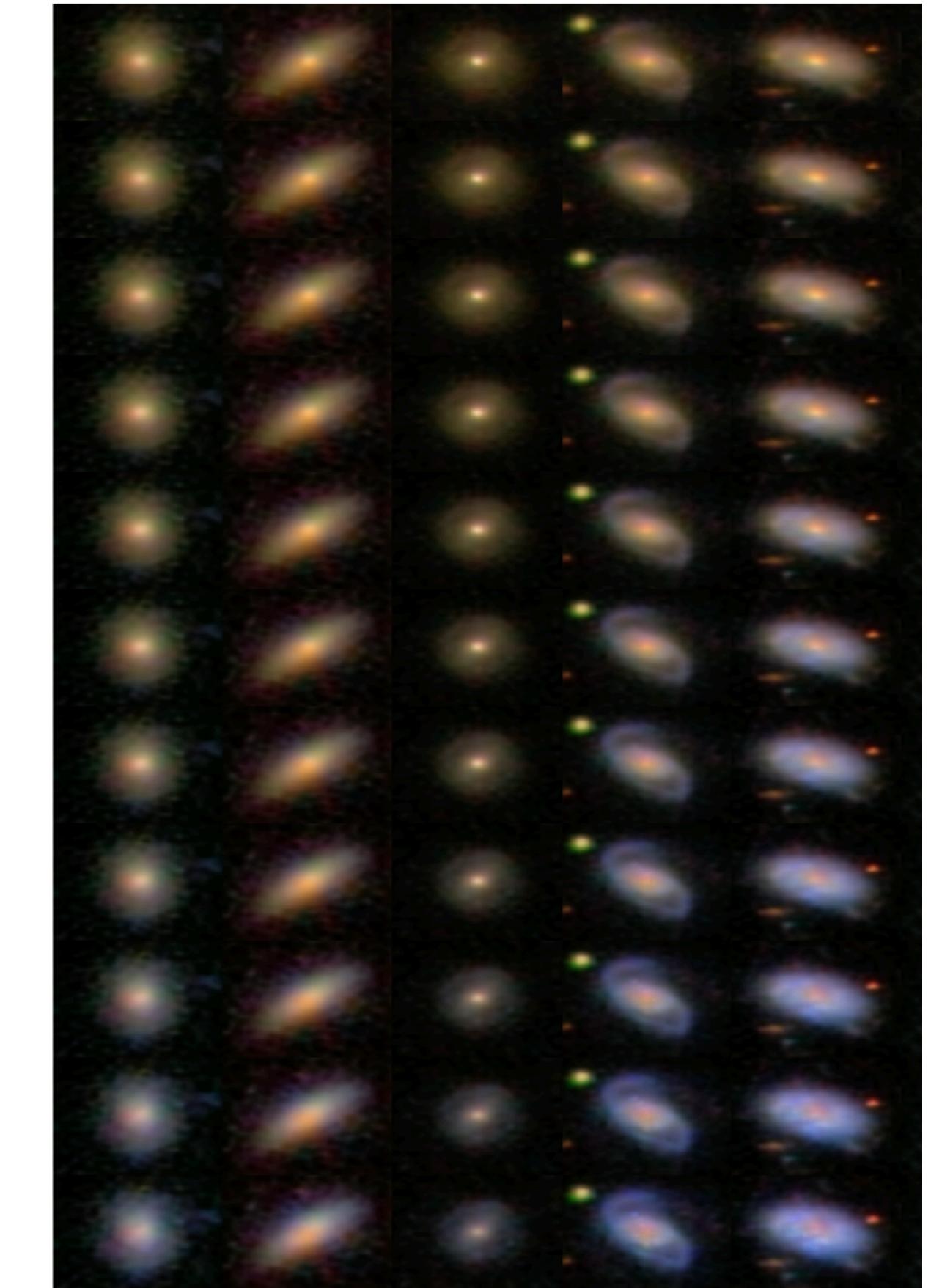


Monte Carlo update

Application of generative models

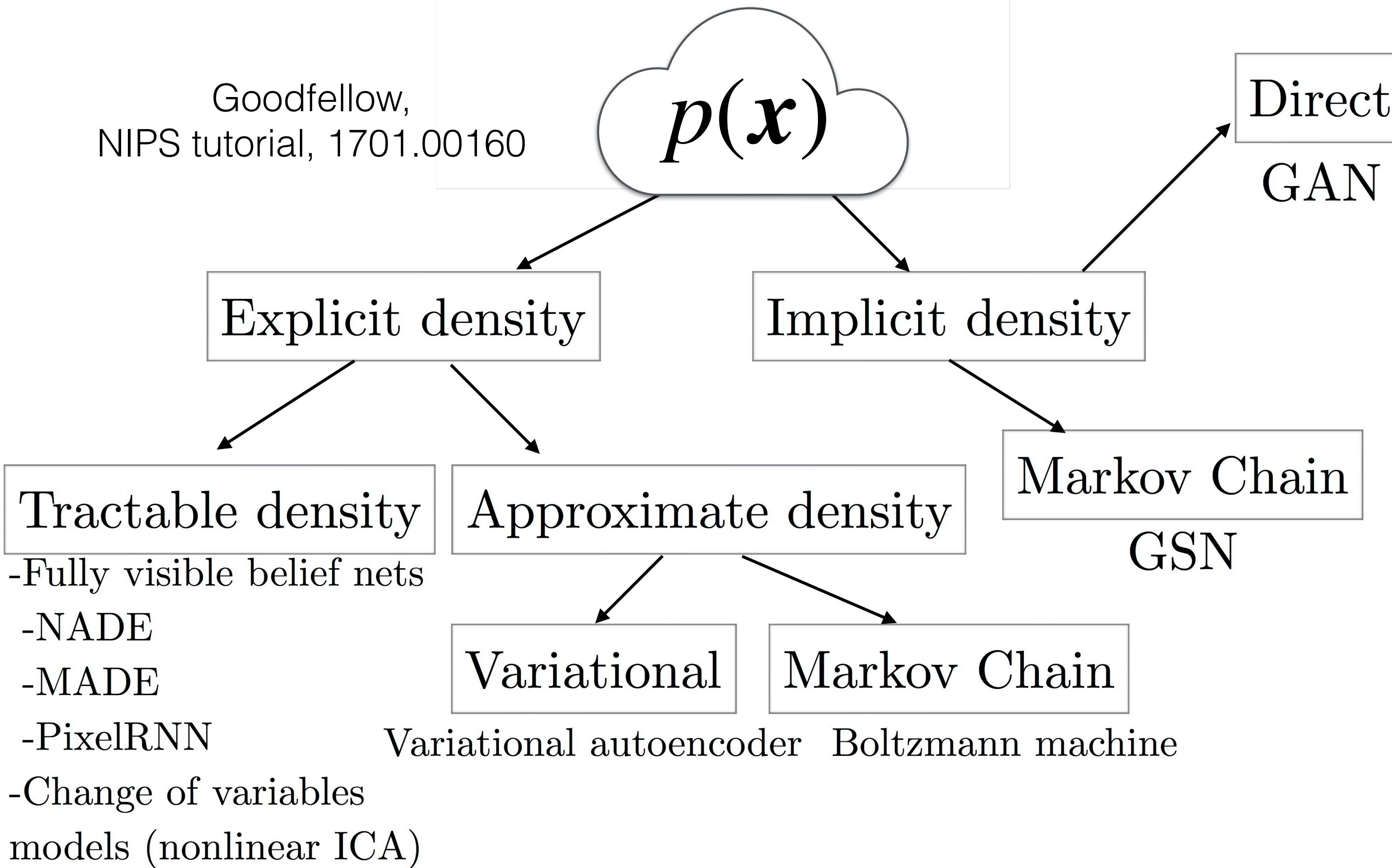


Automatic chemical design,
Gomez-Bombarelli et al, 1610.02415

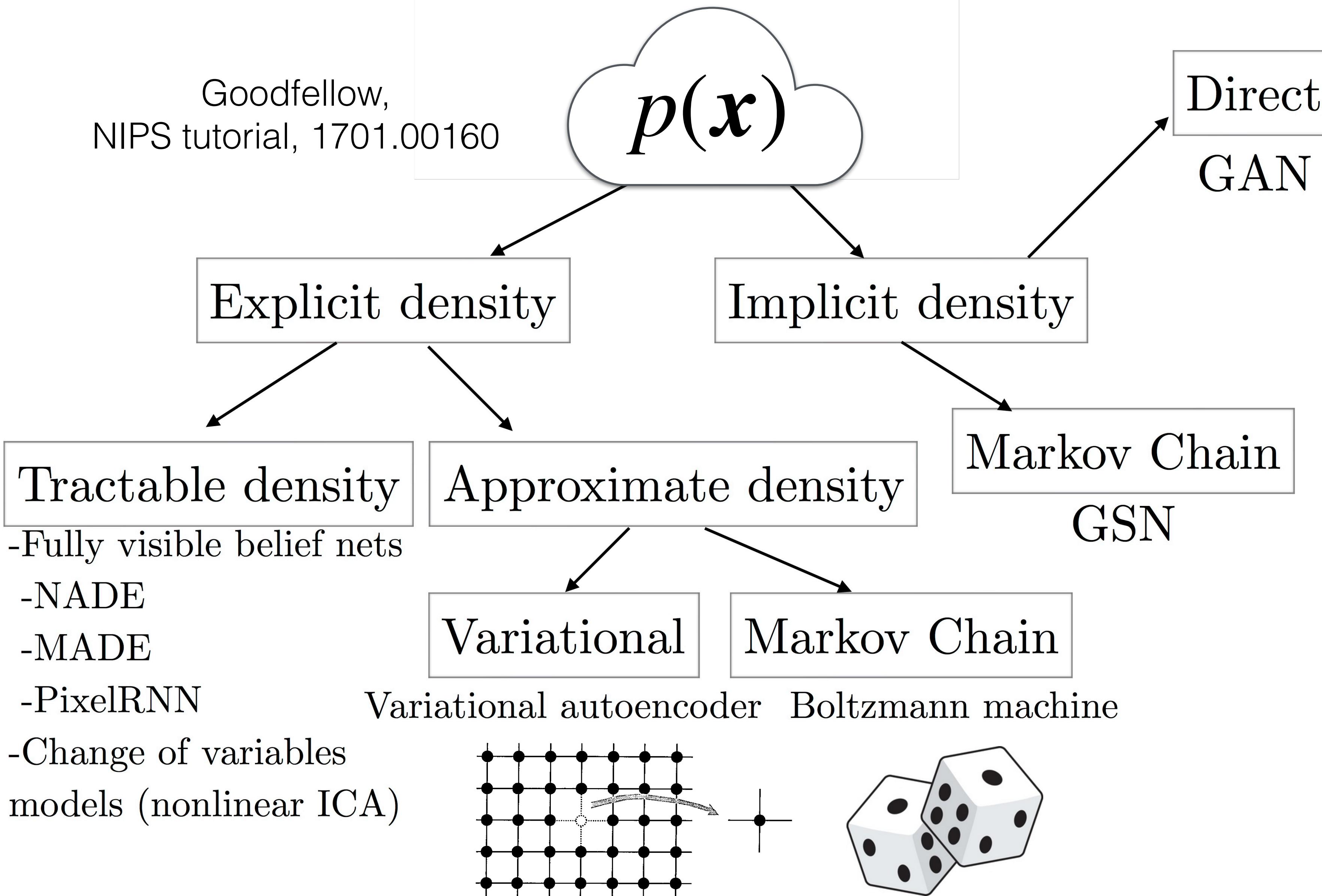


Galaxy evolution
Schawinski et al, unpublished

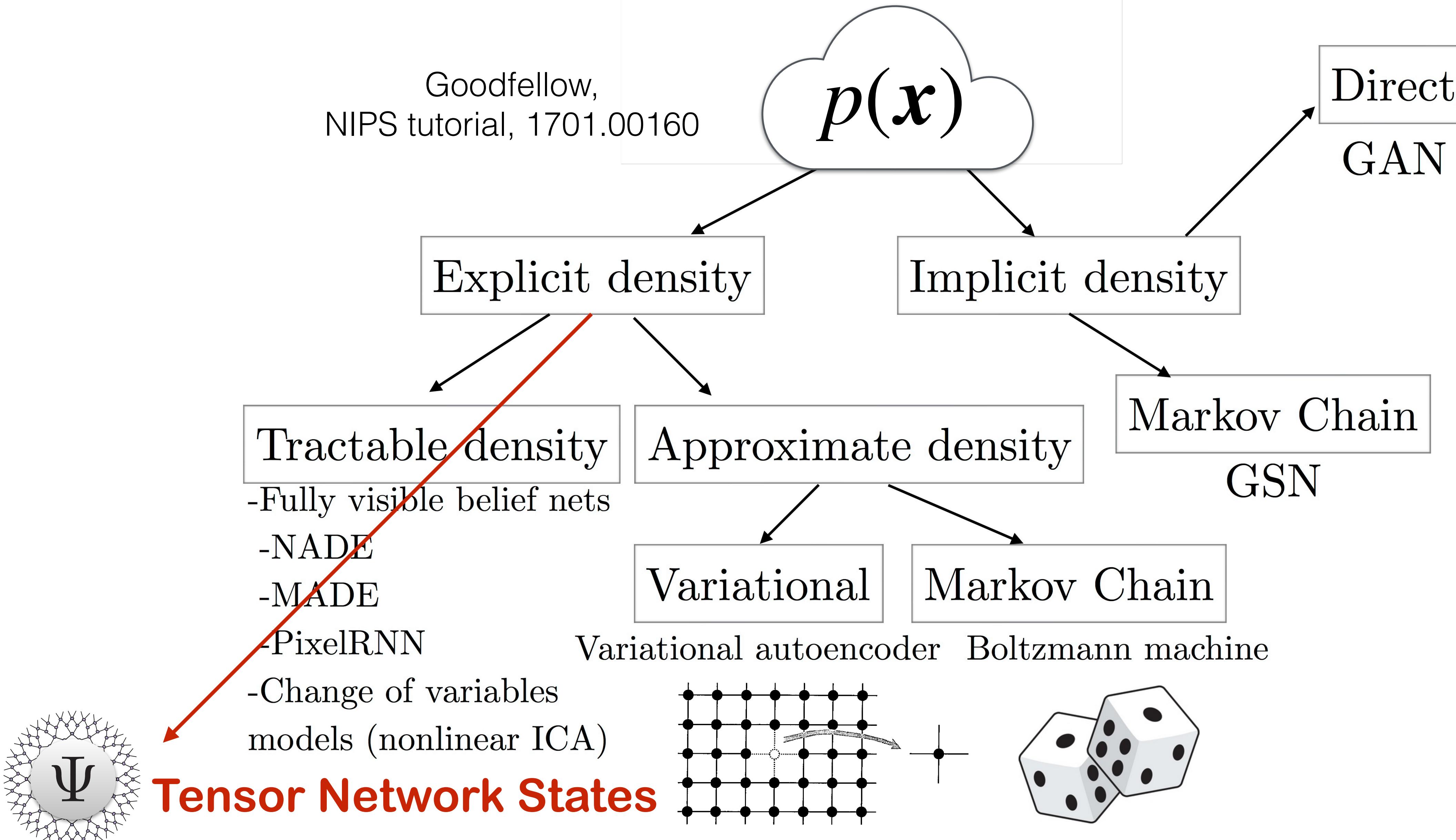
Physics genes of generative models



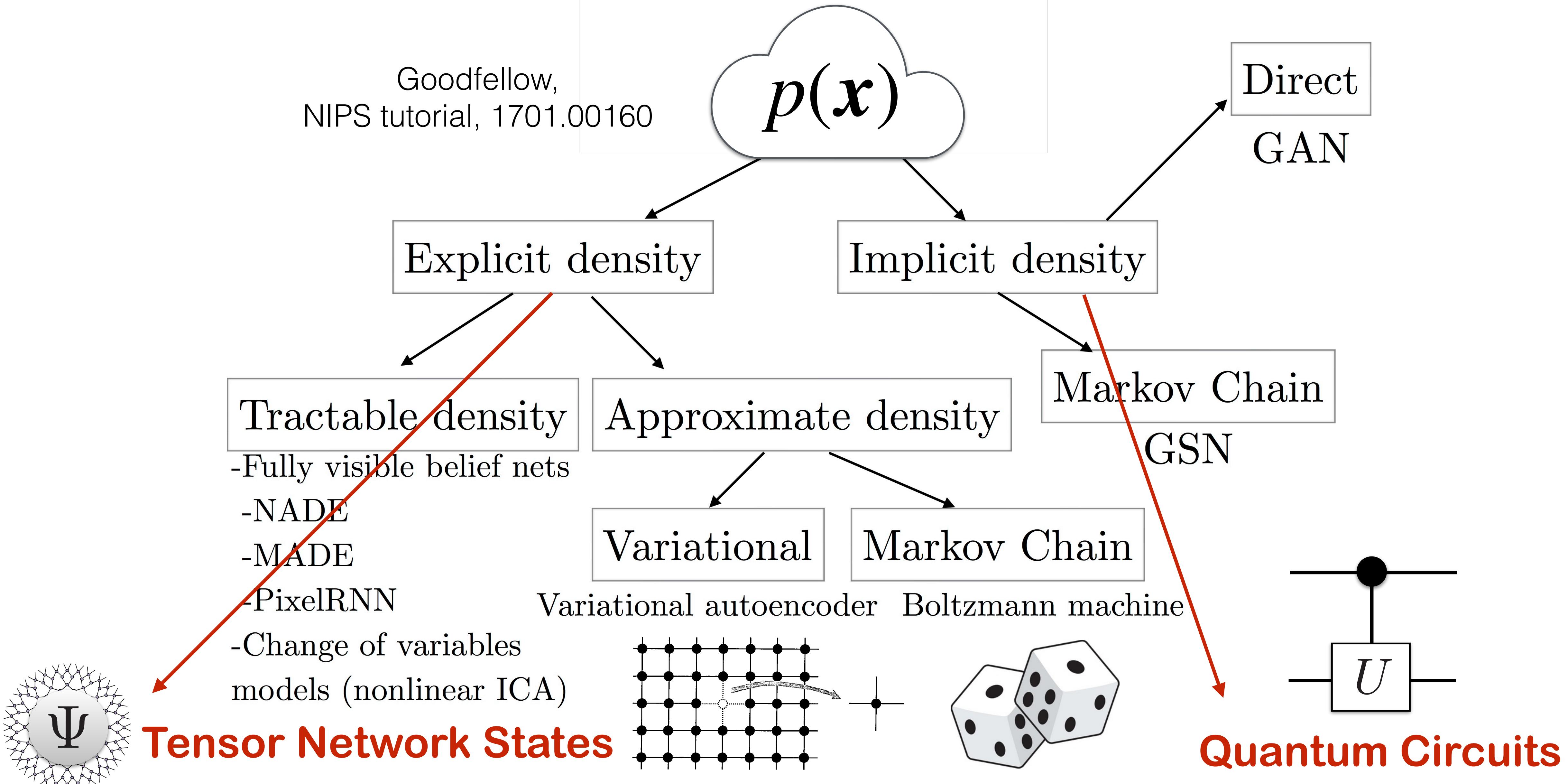
Physics genes of generative models



Physics genes of generative models



Physics genes of generative models



DL as a fluid control problem

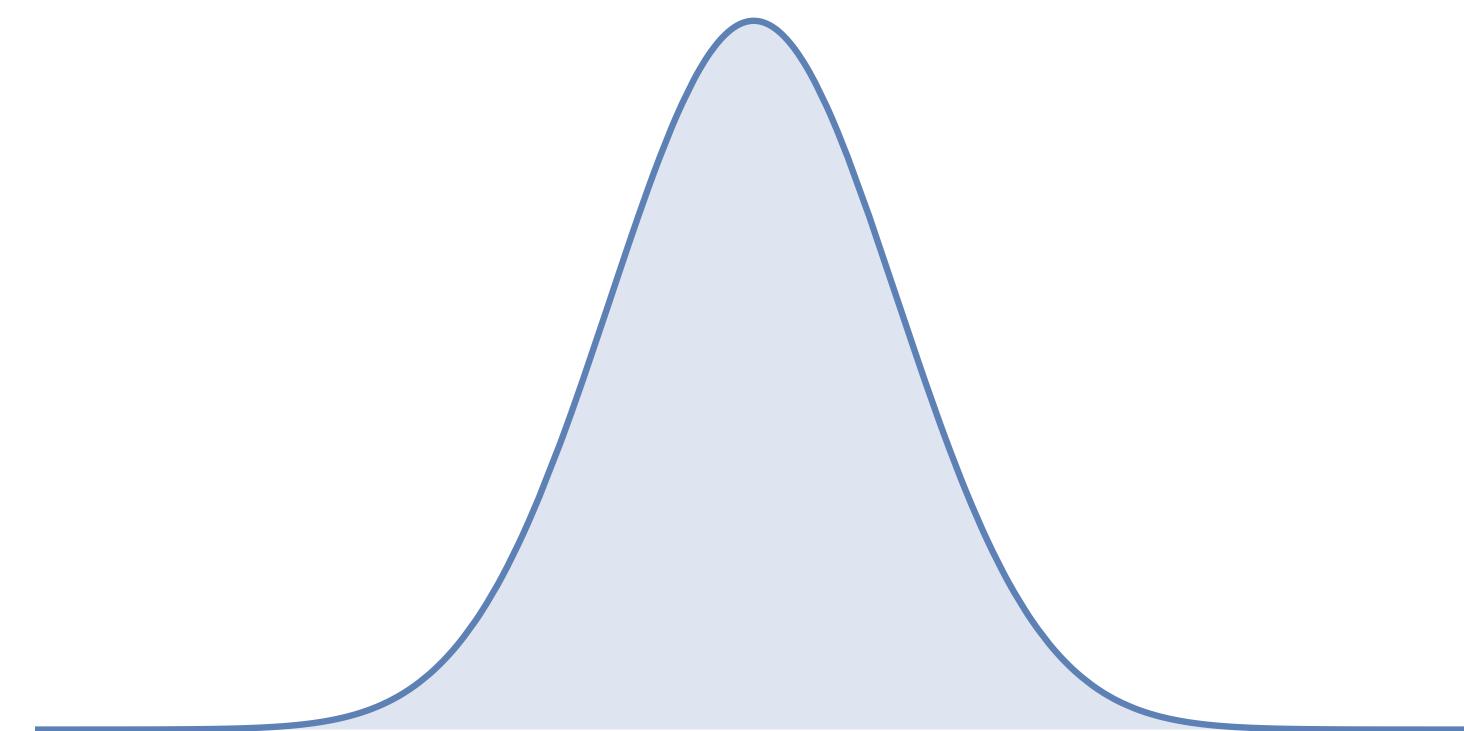
$$\frac{p(z)}{q(\nabla u(z))} = \det \left(\frac{\partial^2 u}{\partial z_i \partial z_j} \right)$$

Monge-Ampère equation
in optimal transport theory

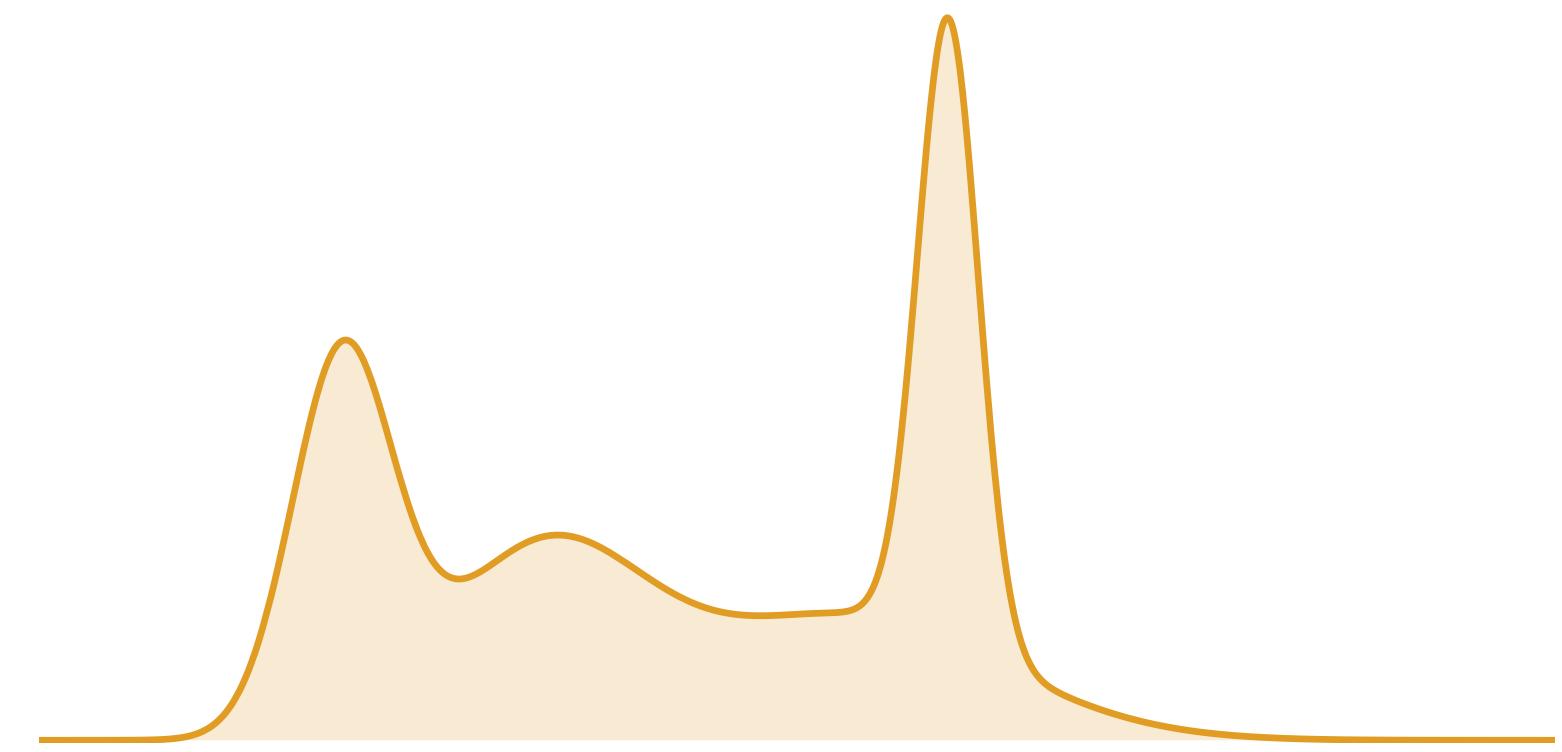
Continuous-time limit
 $\xrightarrow{\epsilon \rightarrow 0}$
 $u(z) = |z|^2/2 + \epsilon \varphi(z)$

$$\frac{\partial p(x, t)}{\partial t} + \nabla \cdot [p(x, t) \nabla \varphi] = 0$$

Continuity equation of
compressible fluids



Simple density



Complex density

DL as a fluid control problem

$$\frac{p(z)}{q(\nabla u(z))} = \det \left(\frac{\partial^2 u}{\partial z_i \partial z_j} \right)$$

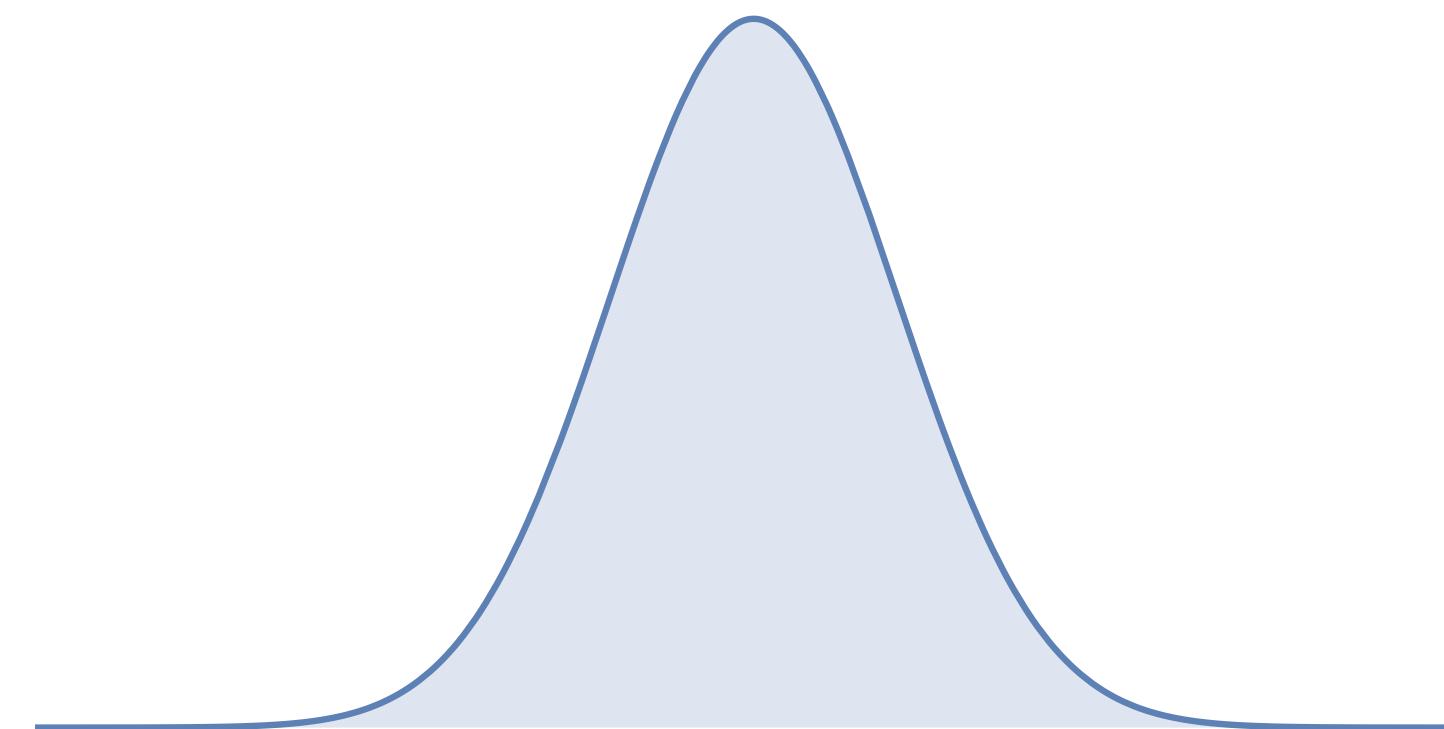
Monge-Ampère equation
in optimal transport theory

Continuous-time limit

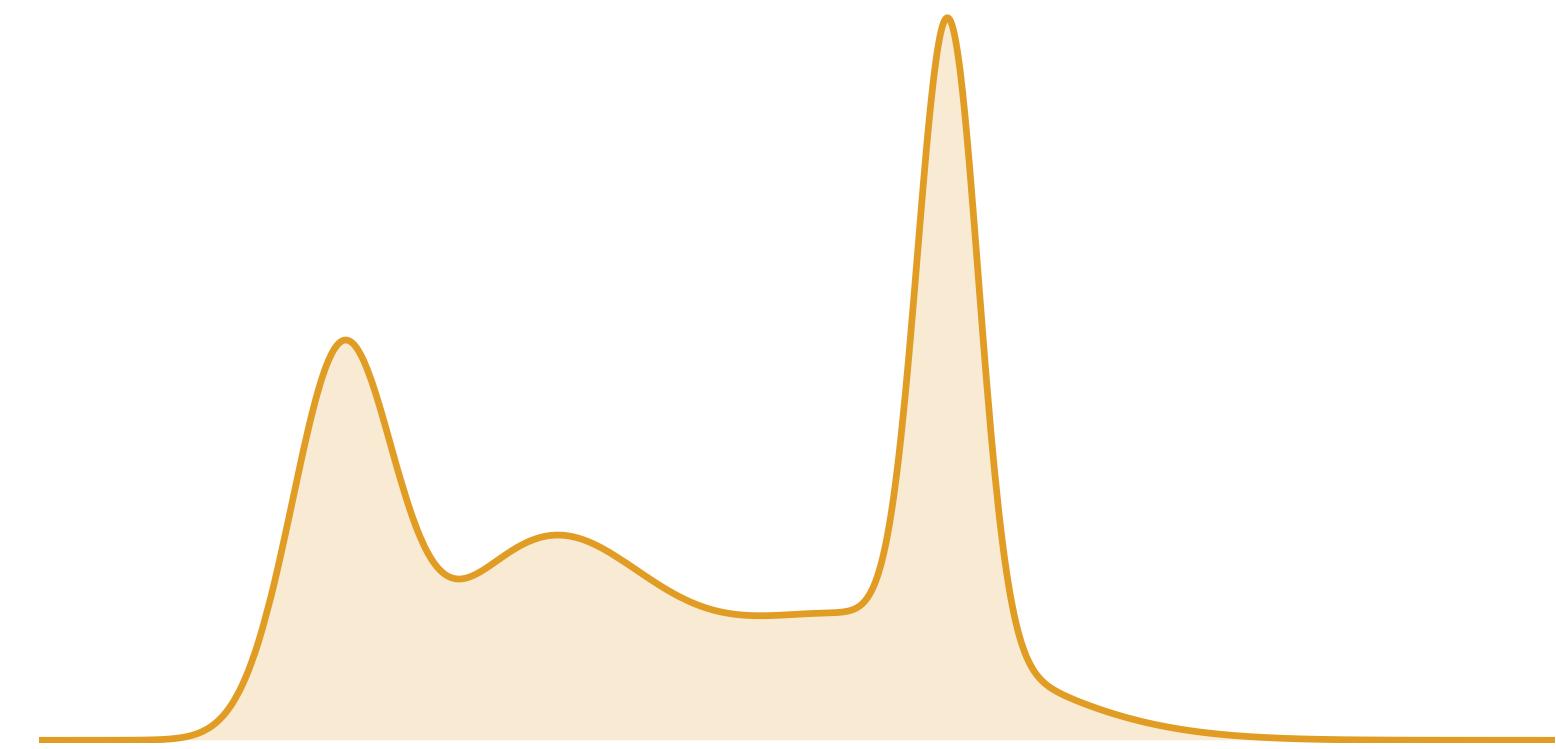
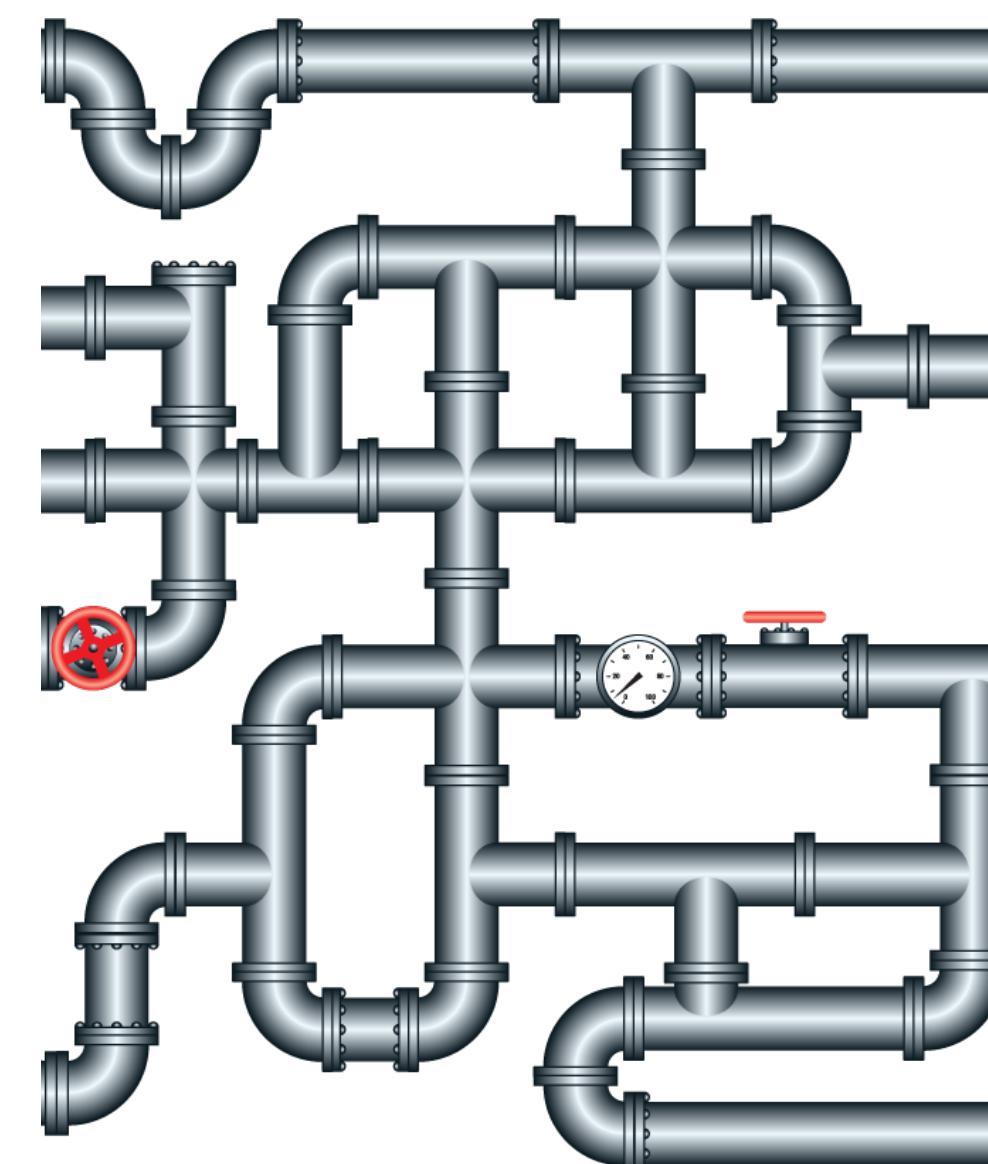
$$u(z) = |z|^2/2 + \epsilon \varphi(z)$$
$$\epsilon \rightarrow 0$$

$$\frac{\partial p(x, t)}{\partial t} + \nabla \cdot [p(x, t) \nabla \varphi] = 0$$

Continuity equation of
compressible fluids



Simple density

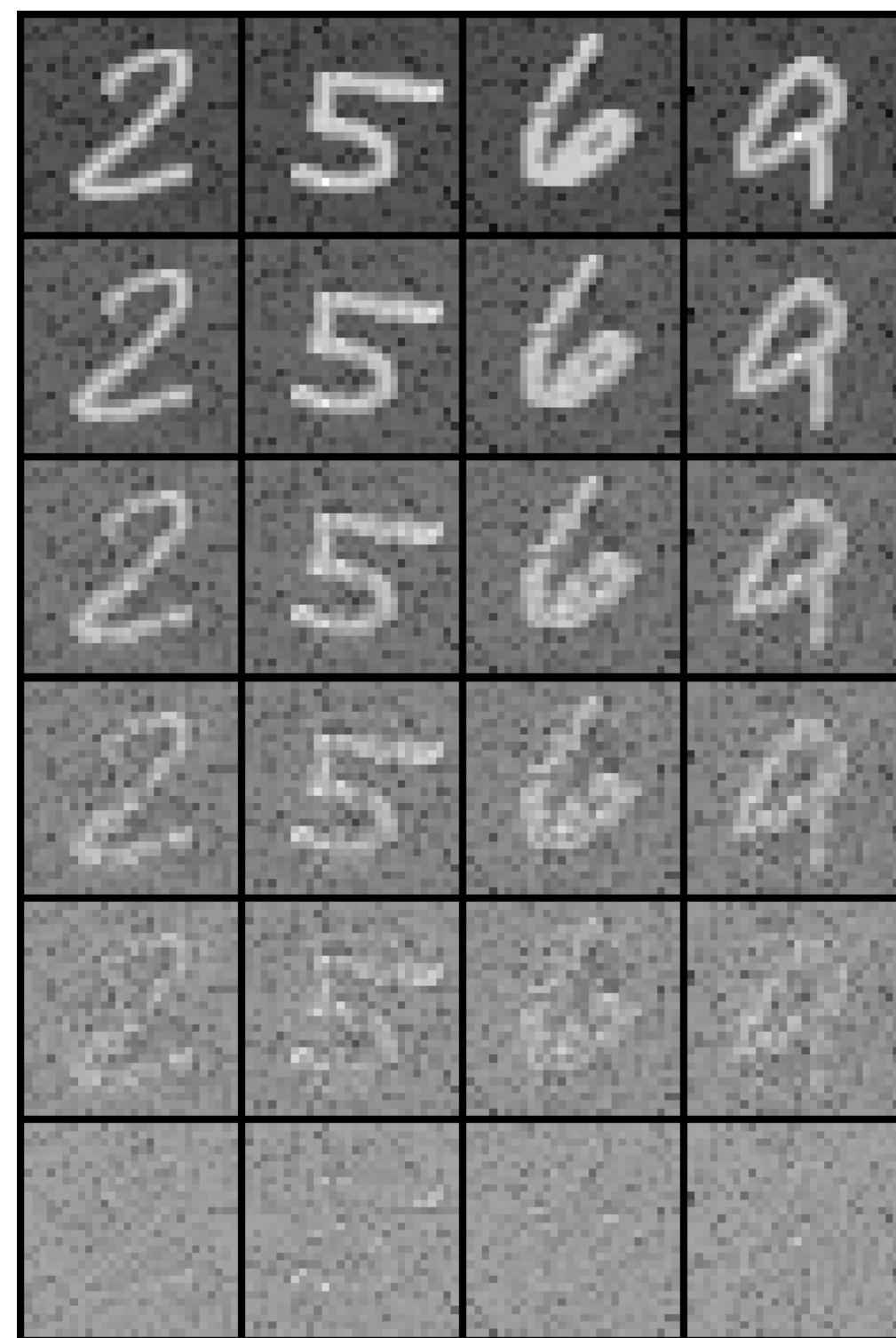


Complex density

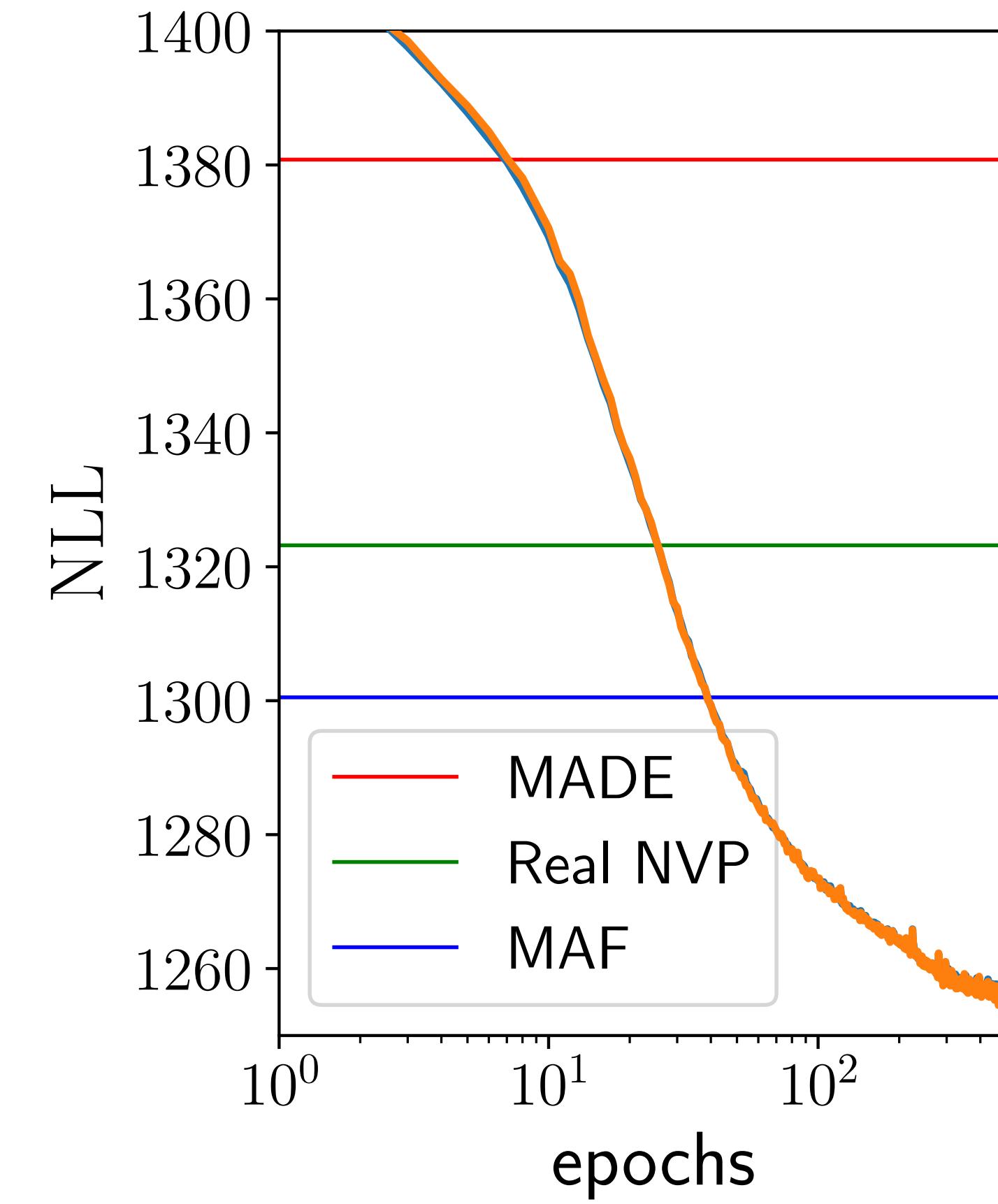
Density estimation of hand-written digits

A standard benchmark for generative models, lower is better

Data space



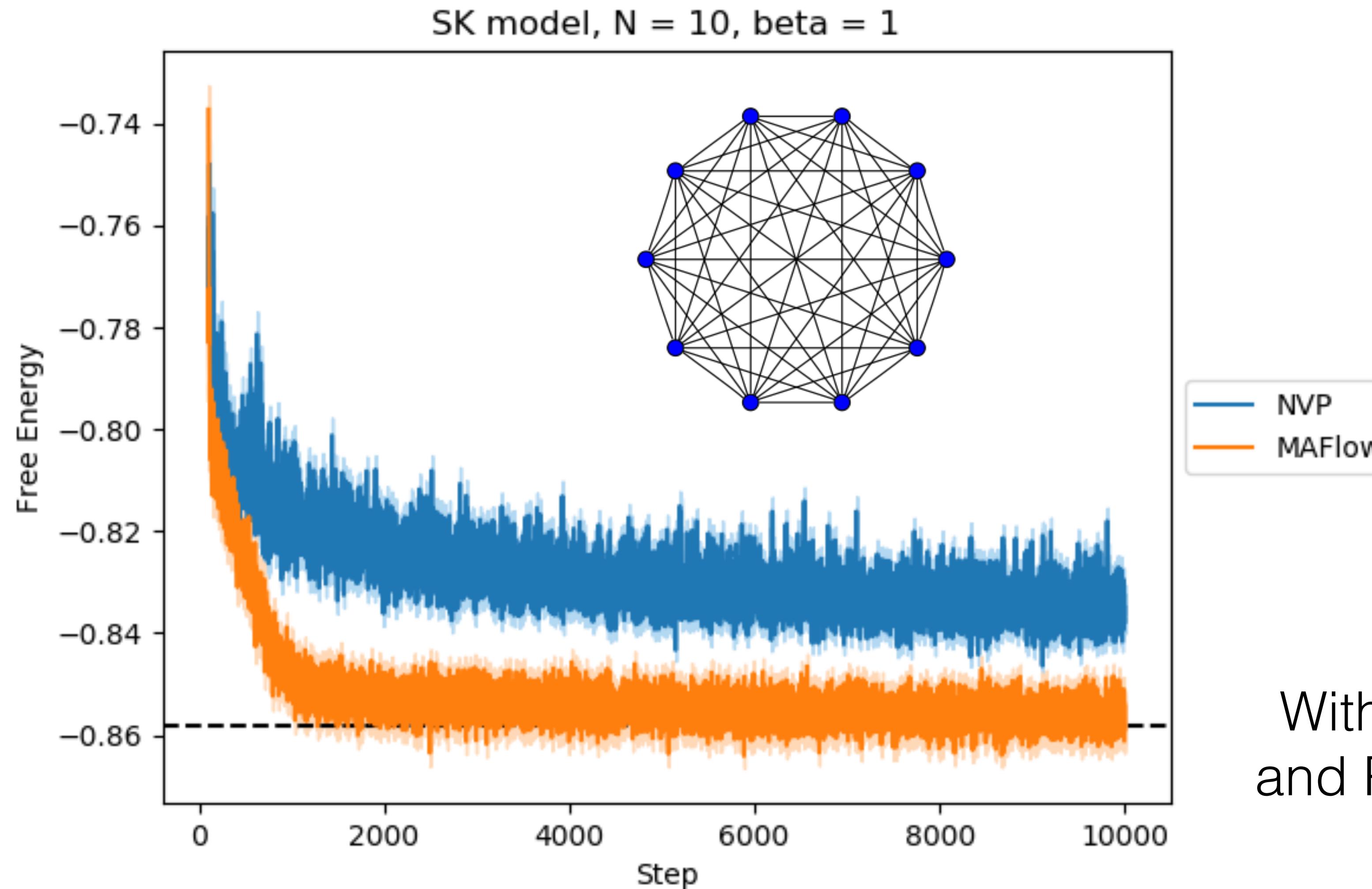
Latent space



'15 '16 '17 '18 Our Result

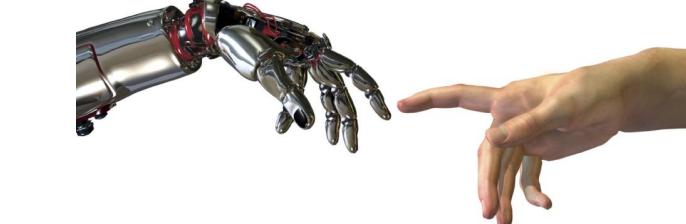
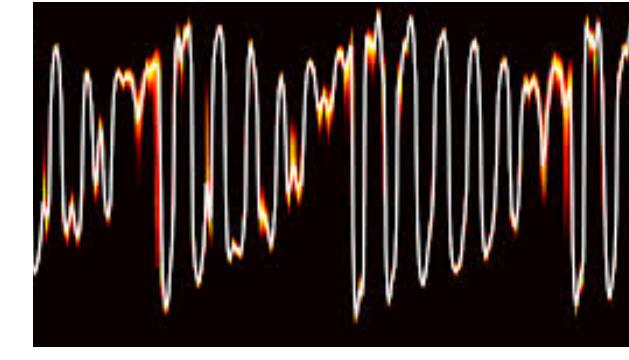
State-of-the-art performance in unstructured density estimation

Variational study of spin glasses

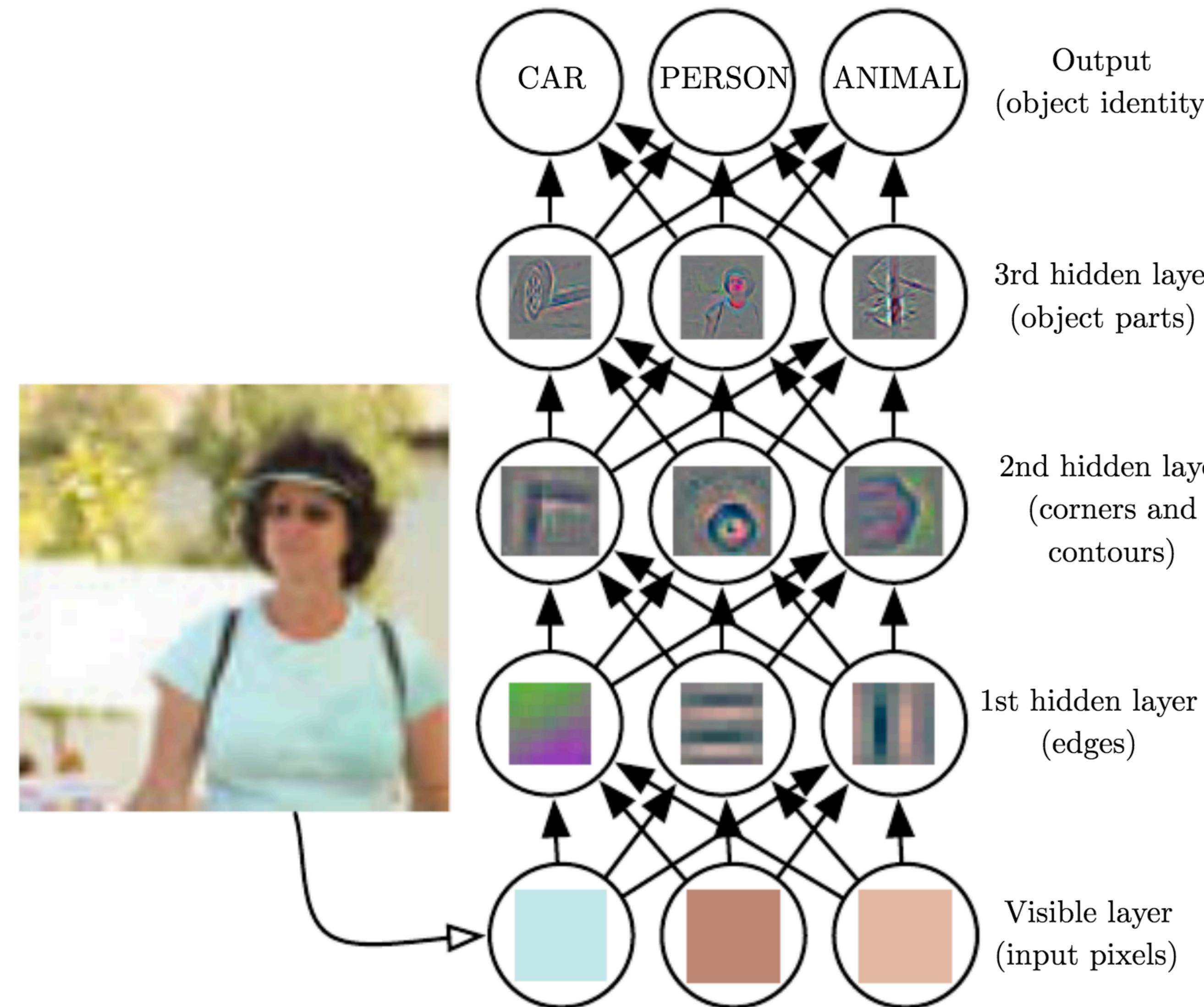


Better variational energy than previous architectures

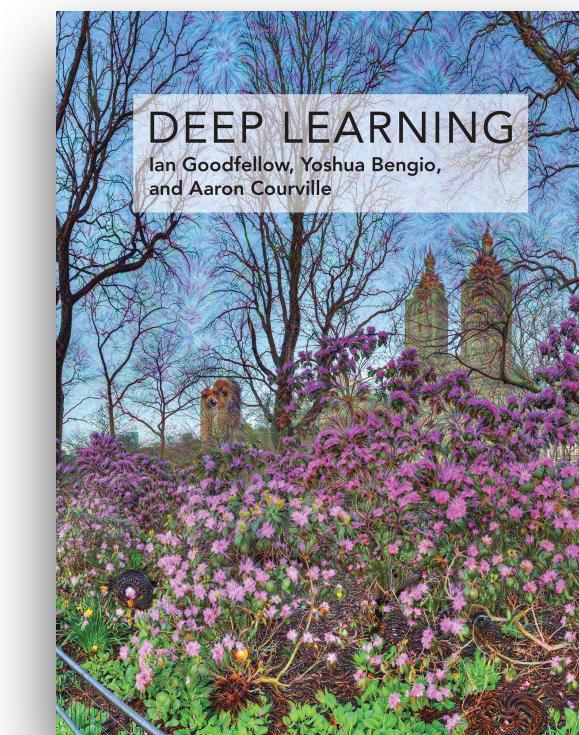
What is the secret behind deep learning?



Representation Learning

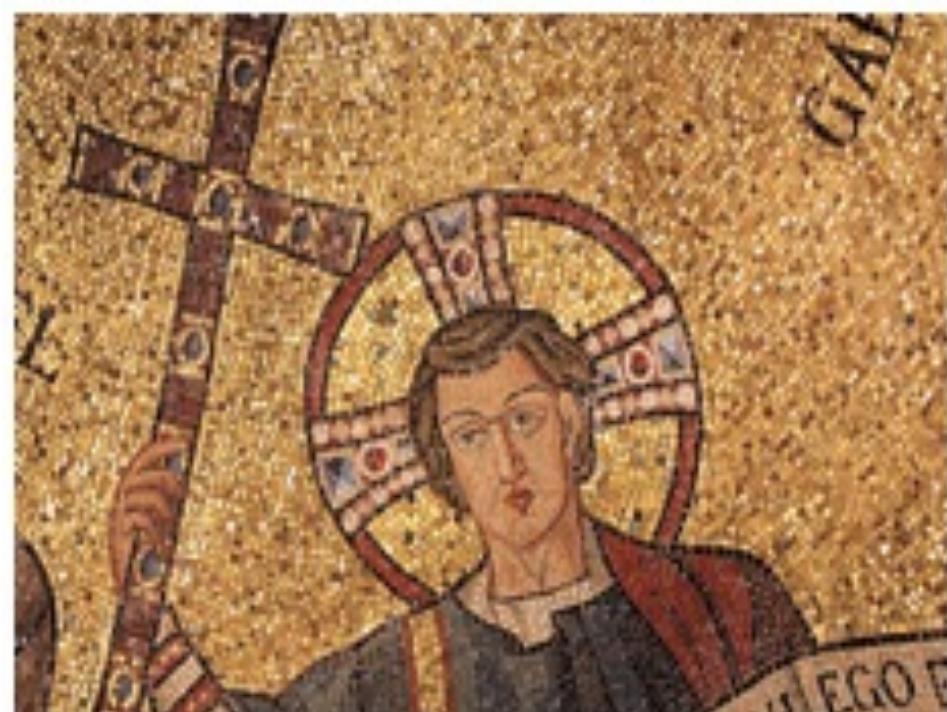


Page 6
Figure 1.2



Magic of learned representations

Neural style transfer



Latent space interpolation



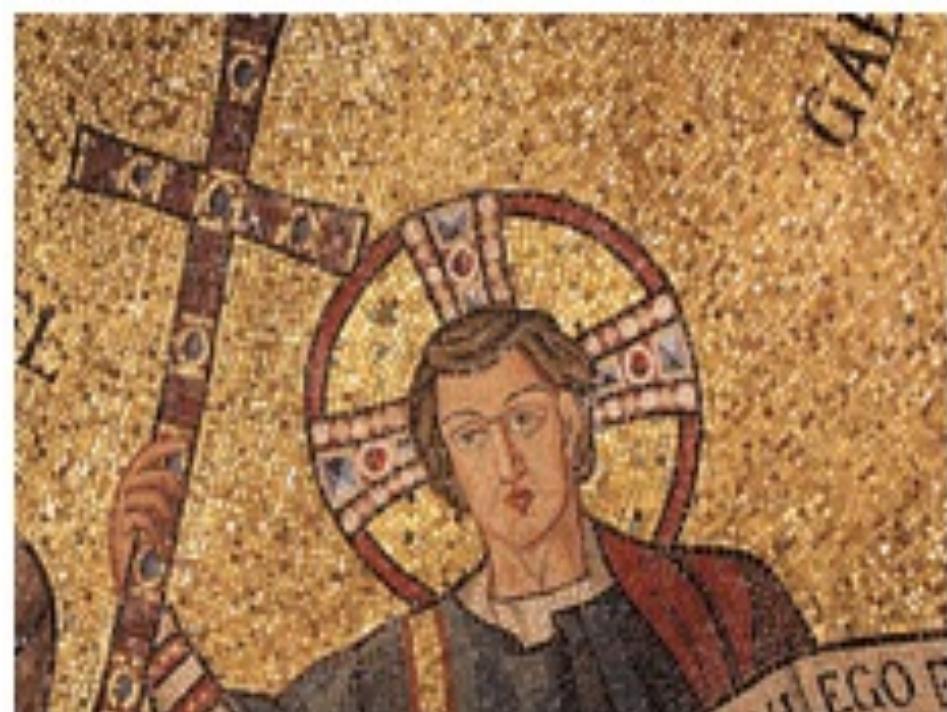
Glow 1807.03039

<https://blog.openai.com/glow/>

Gatys et al, 1508.06576

Magic of learned representations

Neural style transfer



Latent space interpolation



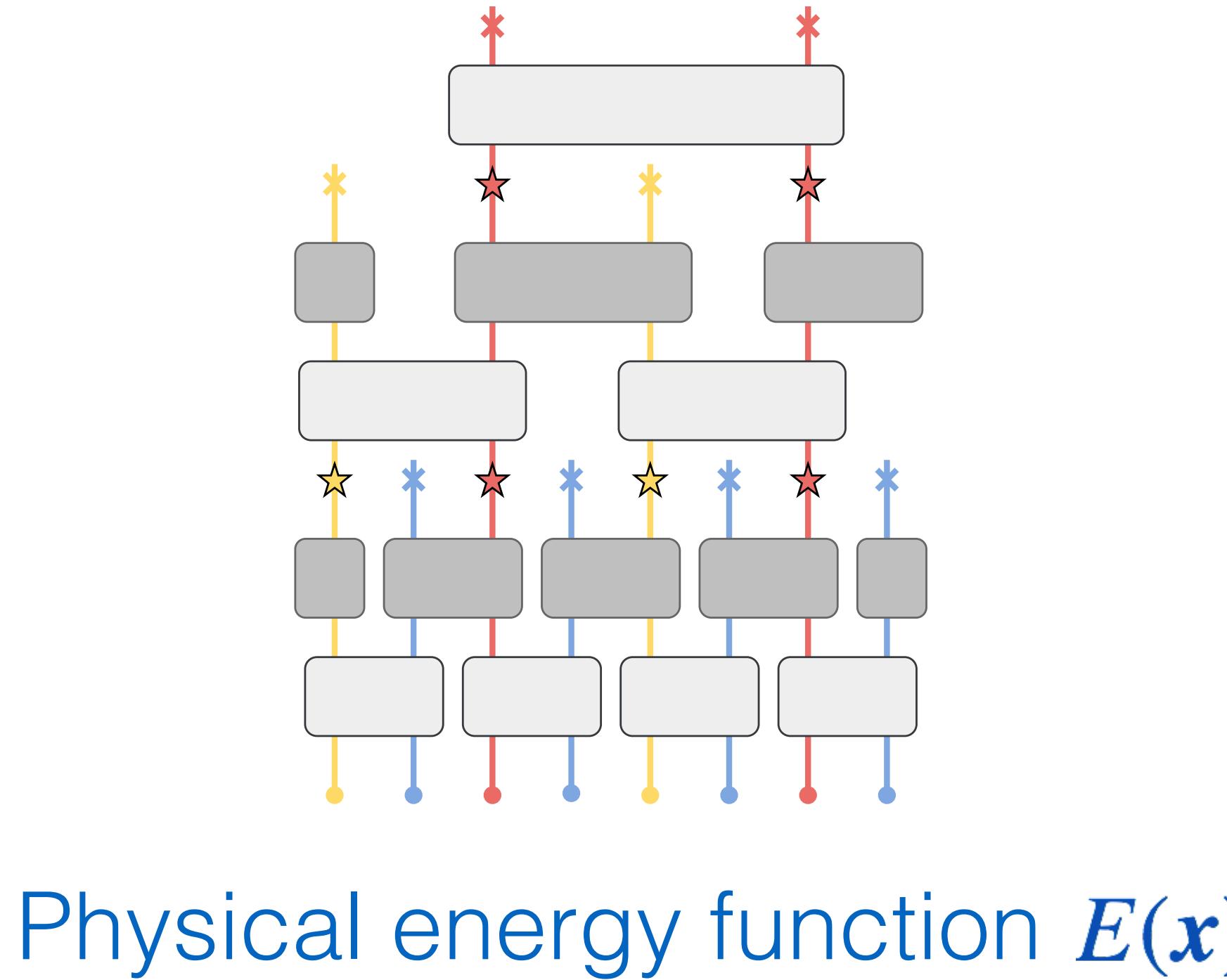
Glow 1807.03039

<https://blog.openai.com/glow/>

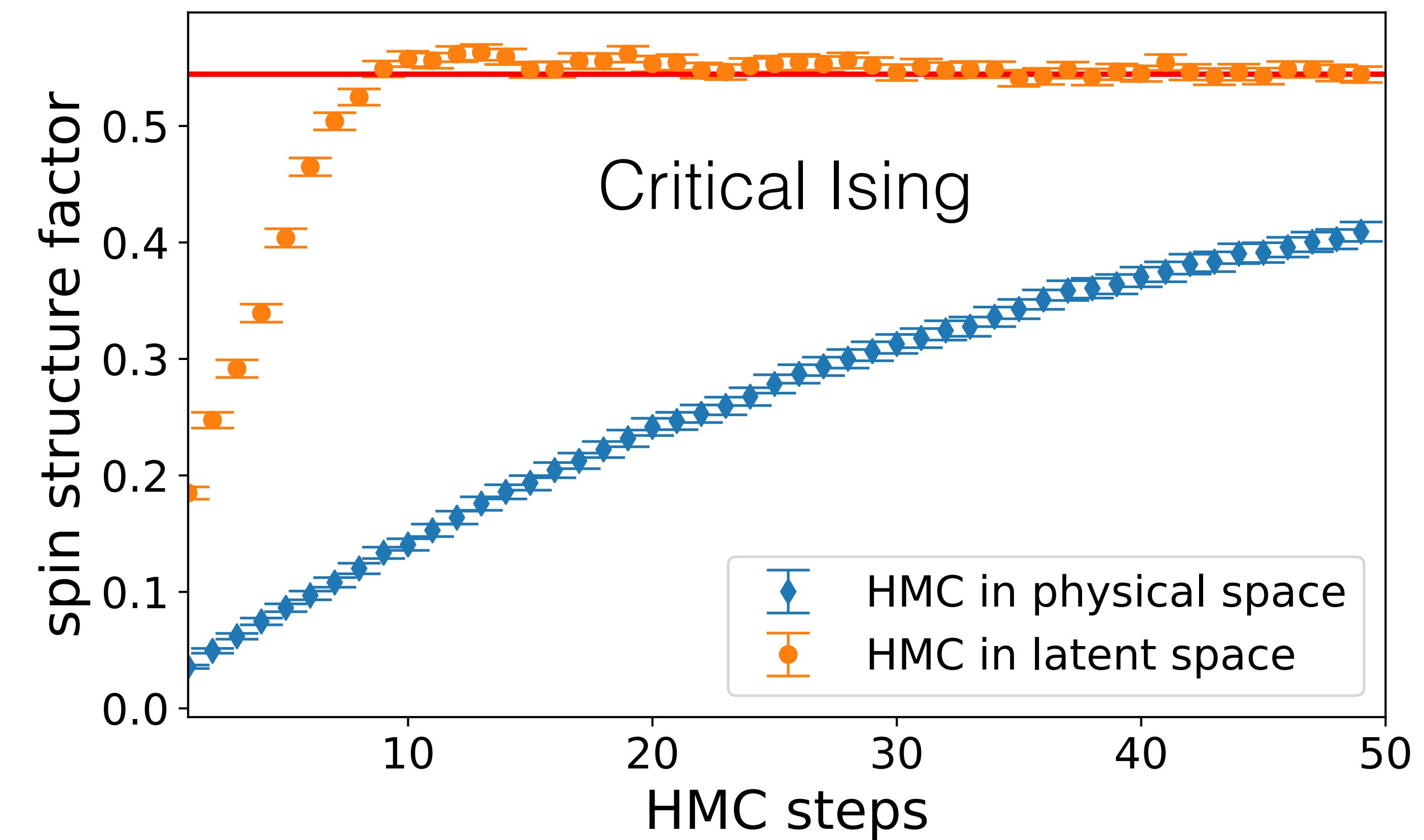
Gatys et al, 1508.06576

Latent space Hybrid MC

Latent space energy function
 $E_{\text{eff}}(z) = E(g(z)) + \ln q(g(z)) - \ln p(z)$



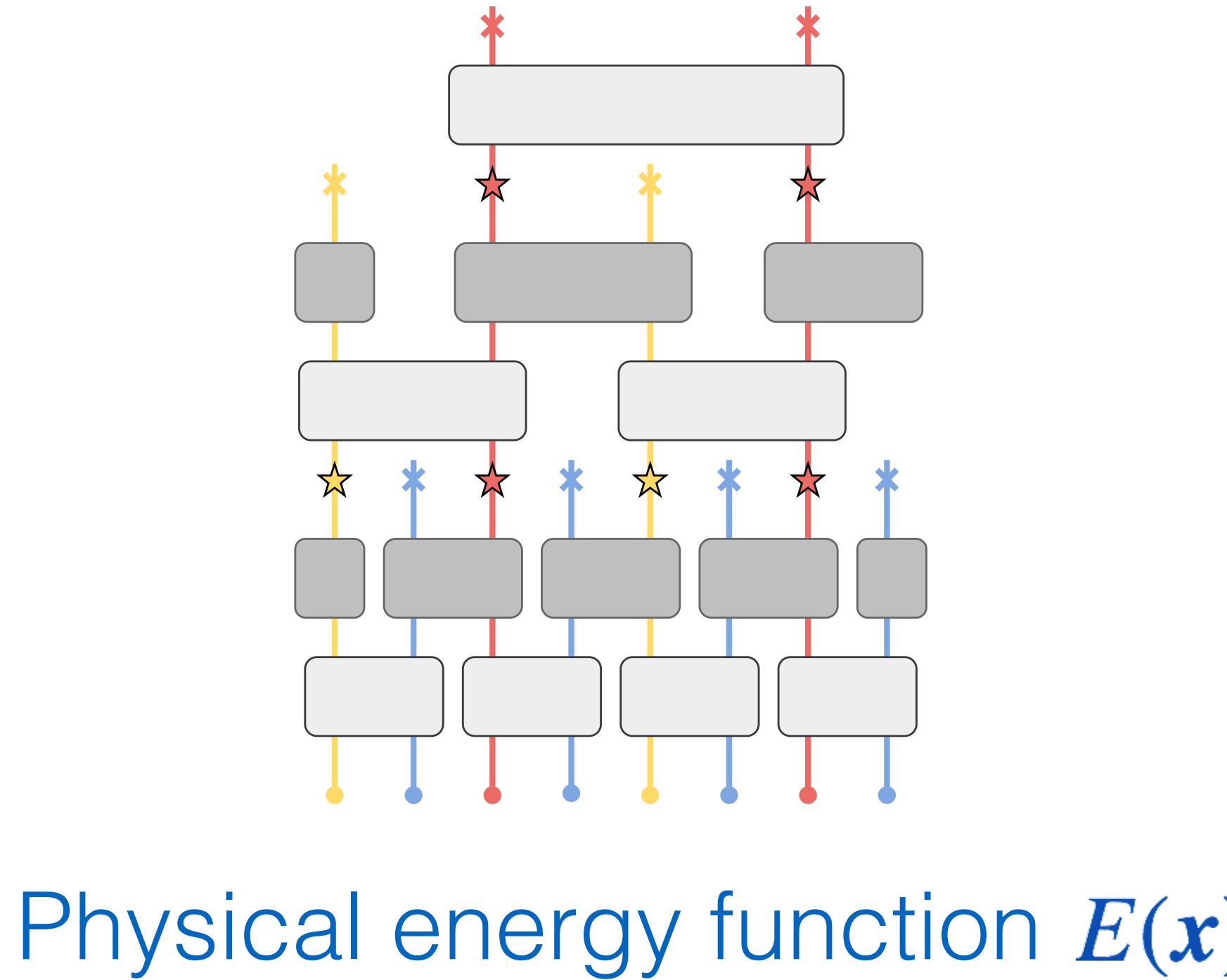
NeuralRG, Shuo-Hui Li and LW, 1802.02840



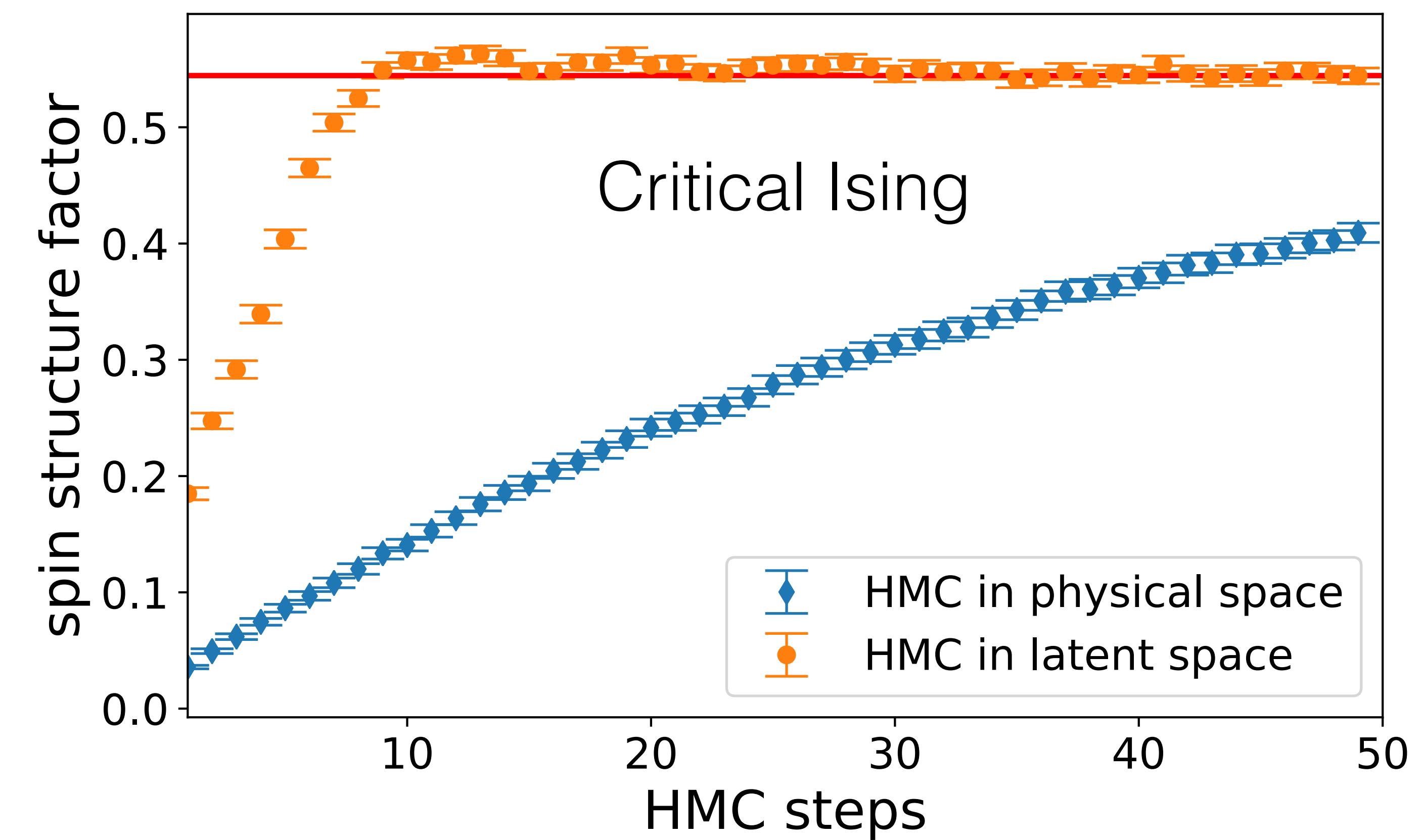
HMC thermalizes faster in the latent space

Latent space Hybrid MC

Latent space energy function
 $E_{\text{eff}}(z) = E(g(z)) + \ln q(g(z)) - \ln p(z)$



NeuralRG, Shuo-Hui Li and LW, 1802.02840



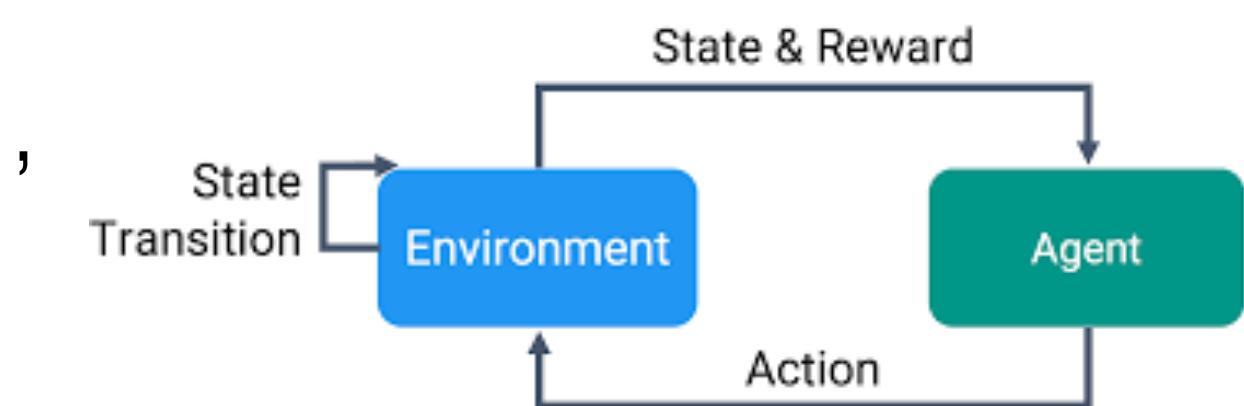
HMC thermalizes faster in the latent space

Remarks on accelerated MC

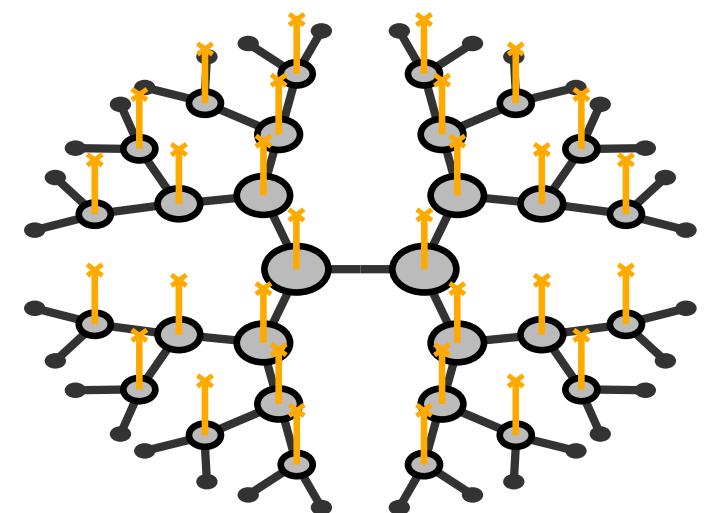
1. Cheap **surrogate function** for MC weight Neal 96' Jun. S Liu 01' **A recommender engine** for MC updates when the surrogate is a generative model: Huang, LW, 1610.02746, Liu, Qi, Meng, Fu, 1610.03137



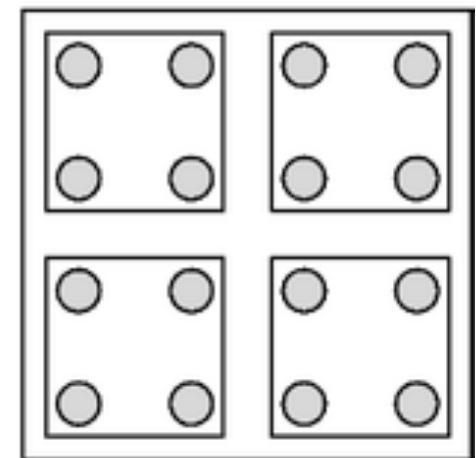
2. Reinforcement learning the **transition kernel**: Song et al, 1706.07561, Levy et al 1711.09268, Cusumano-Towner et al 1801.03612, Bojesen, 1808.09095



3. Performs MC in the **variationally learned disentangled representation**: Wavelet MC, Ismail 03' , NeuralRG 18'

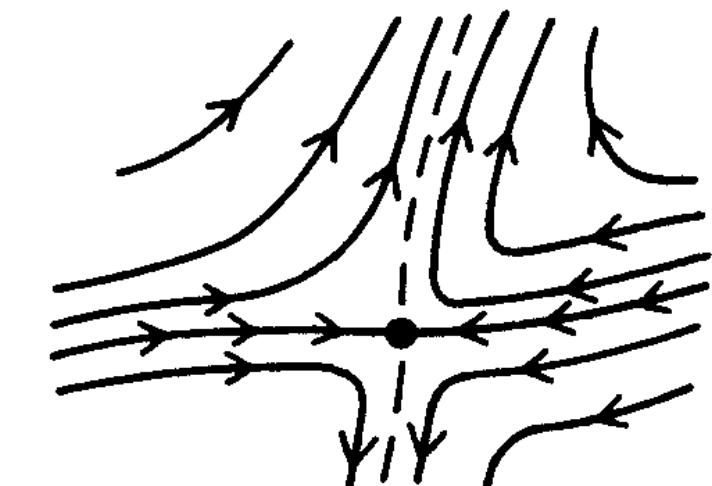


Deep learning and RG

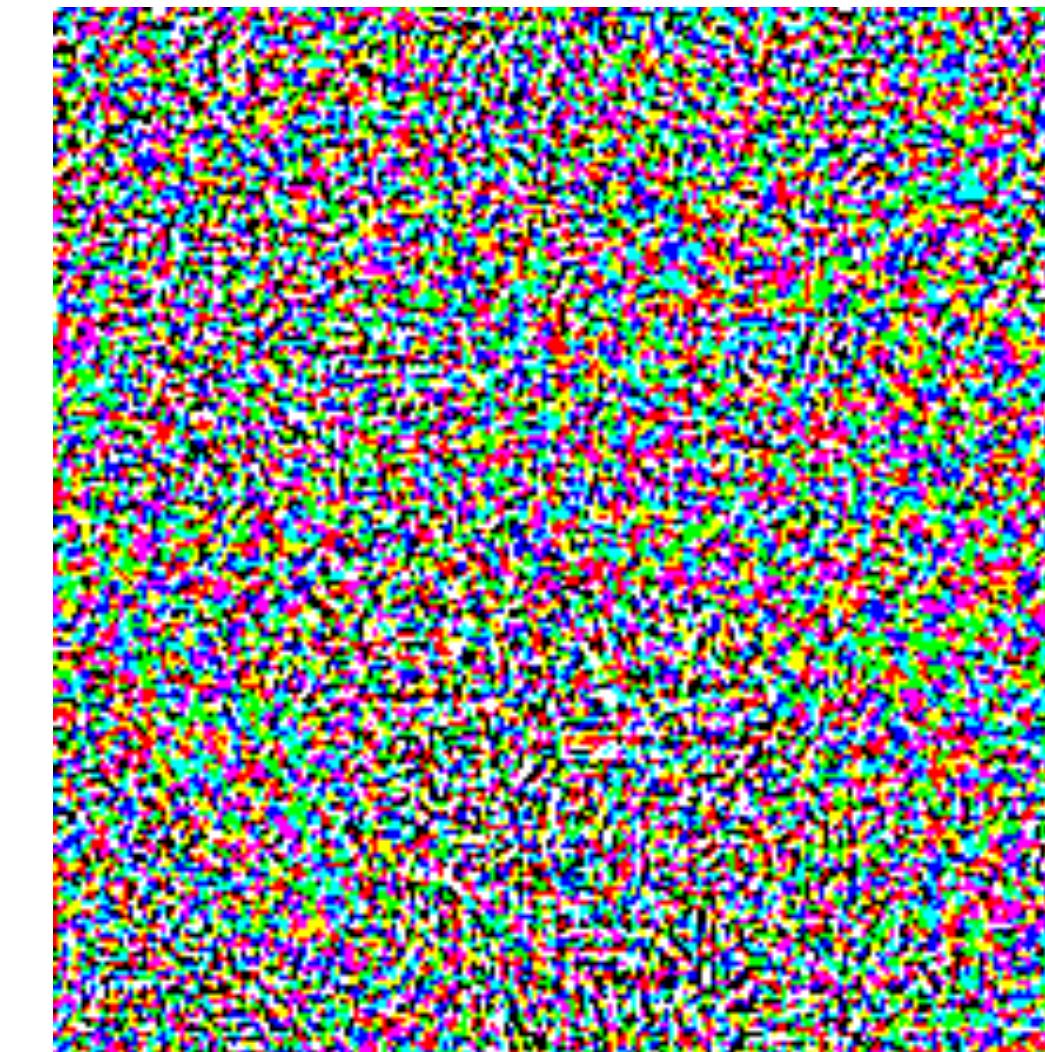


't Hooft, Gross, Wilczek, Kadanoff, Wilson, Fisher...

Bény, Mehta, Schwab, Lin, Tegmark, You, Qi ...



+ .007 ×



=



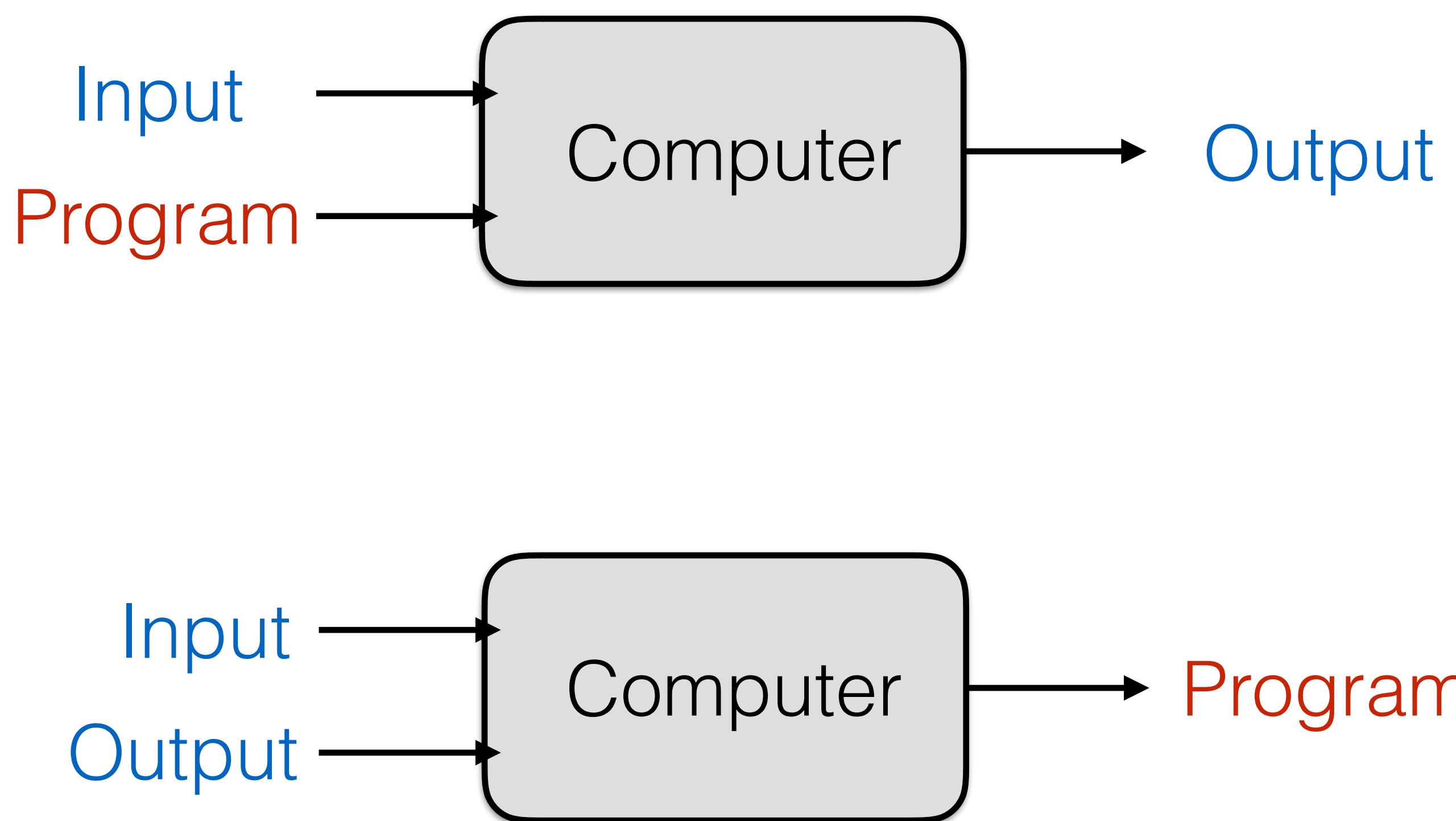
Panda
Confidence 58%

Goodfellow et al, 2014

Gibbon
Confidence 99%

Vulnerability of deep learning, Kenway, 1803.06111 & 1803.10995

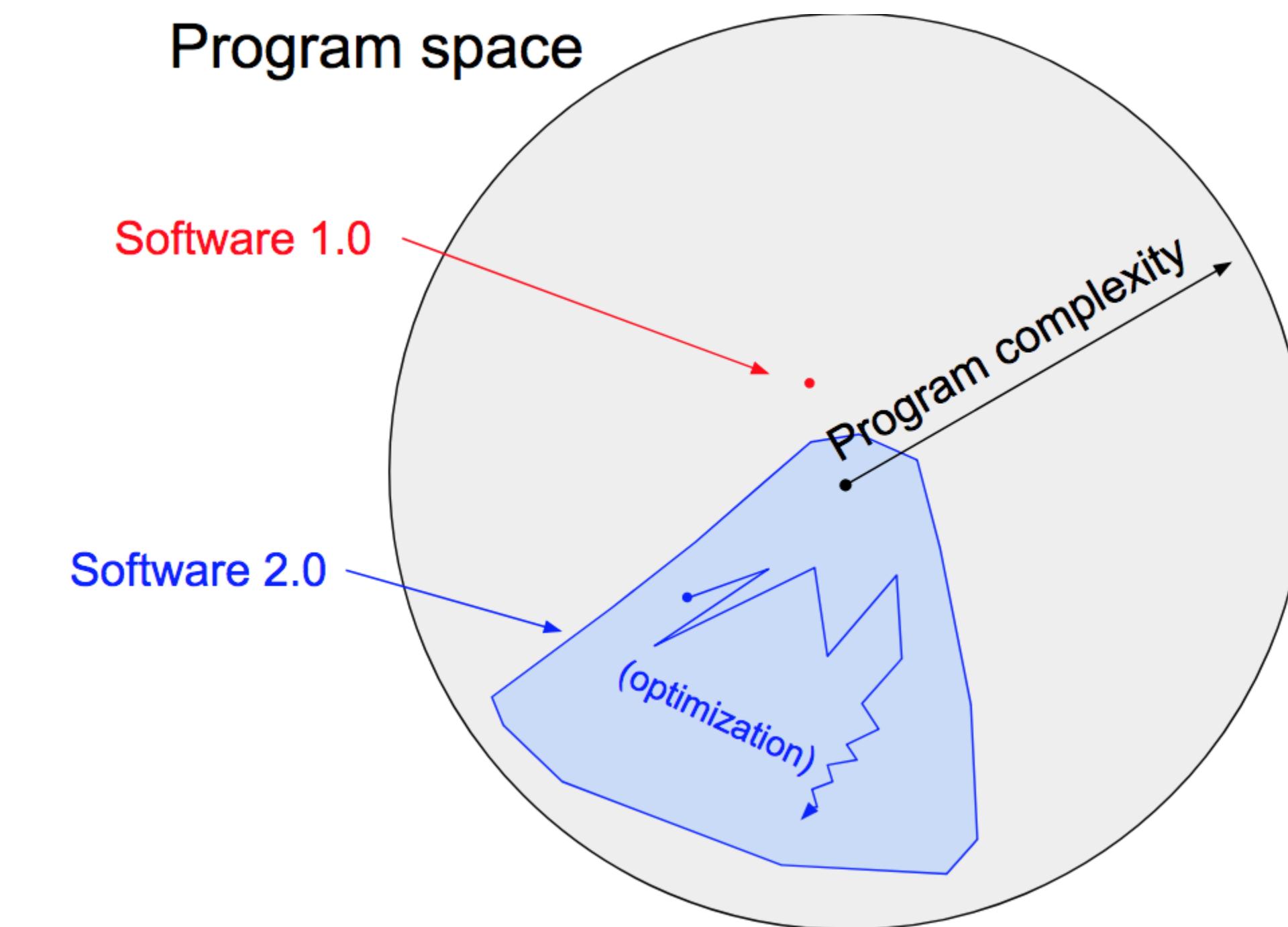
Differentiable Programming



Andrej Karpathy

Director of AI at Tesla. Previously Research Scientist at OpenAI and PhD student at Stanford. I like to train deep neural nets on large datasets.

<https://medium.com/@karpathy/software-2-0-a64152b37c35>



Writing software 2.0 by searching in the program space

Differentiable Programming

Benefits compared to 1.0

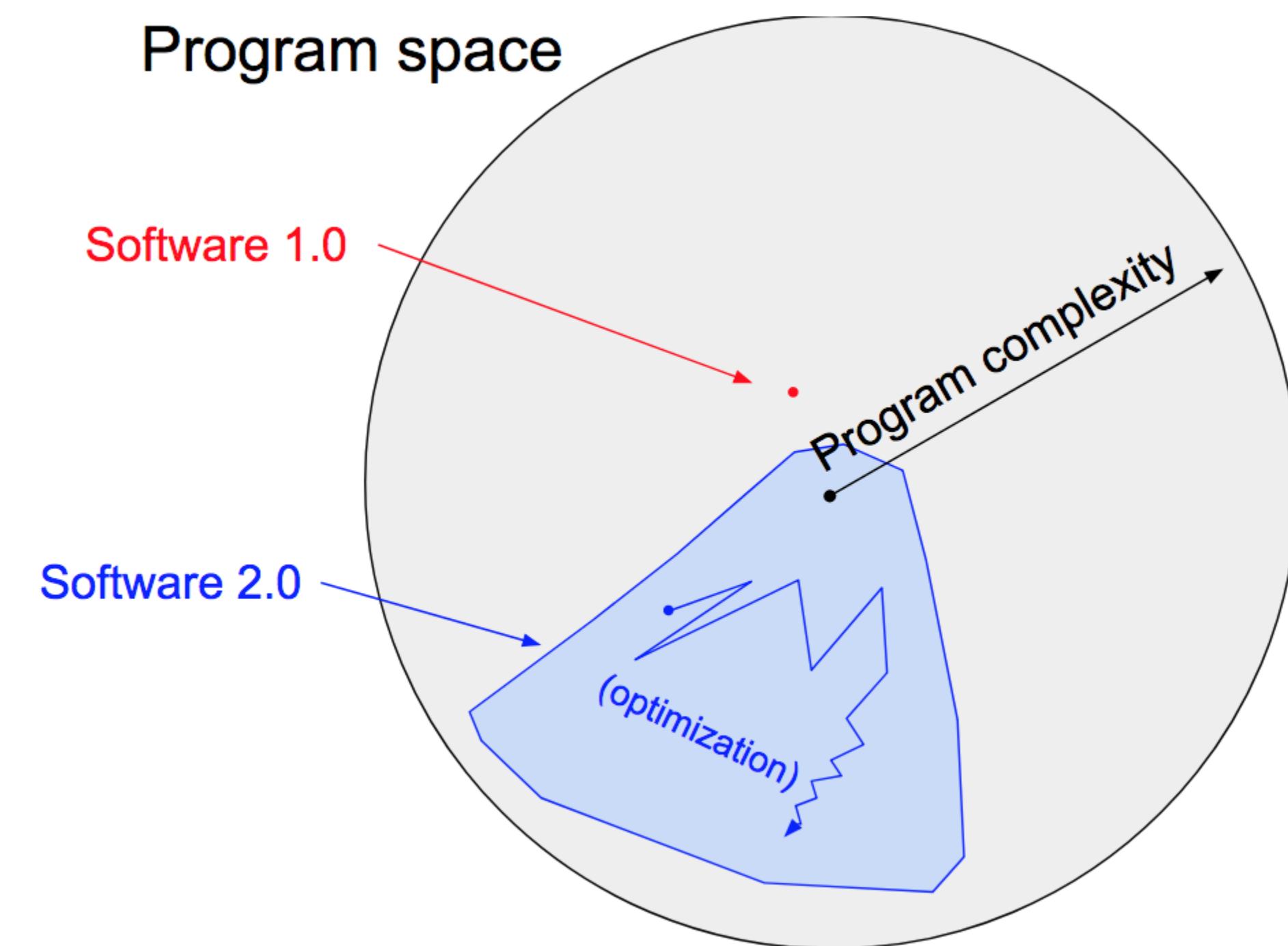
- Computationally homogeneous
- Simple to bake into silicon
- Constant running time
- Constant memory usage
- Highly portable & agile
- Modules can meld into an optimal whole
- Better than humans



Andrej Karpathy

Director of AI at Tesla. Previously Research Scientist at OpenAI and PhD student at Stanford. I like to train deep neural nets on large datasets.

<https://medium.com/@karpathy/software-2-0-a64152b37c35>



Writing software 2.0 by searching in the program space

Differentiable Scientific Programming

- Most linear algebra libraries are [differentiable](#)
- Condition/Sort/Permutations are also differentiable
- ODE integrators are differentiable with [O\(1\) memory](#)
- [Differentiable ray tracer](#) and [Differentiable fluid simulations](#)
- Differentiable Monte Carlo/Tensor Network/Functional RG/
Dynamical Mean Field Theory/Density Functional Theory...

Differentiable Eigensolver

$$\textcolor{red}{H}\Psi = \textcolor{blue}{\Psi}\Lambda$$

Forward mode: What happens if $\textcolor{red}{H} = H + dH$? Perturbation theory

Reverse mode: How should I change H given
 $d\Psi$ and $d\Lambda$? Inverse perturbation theory!

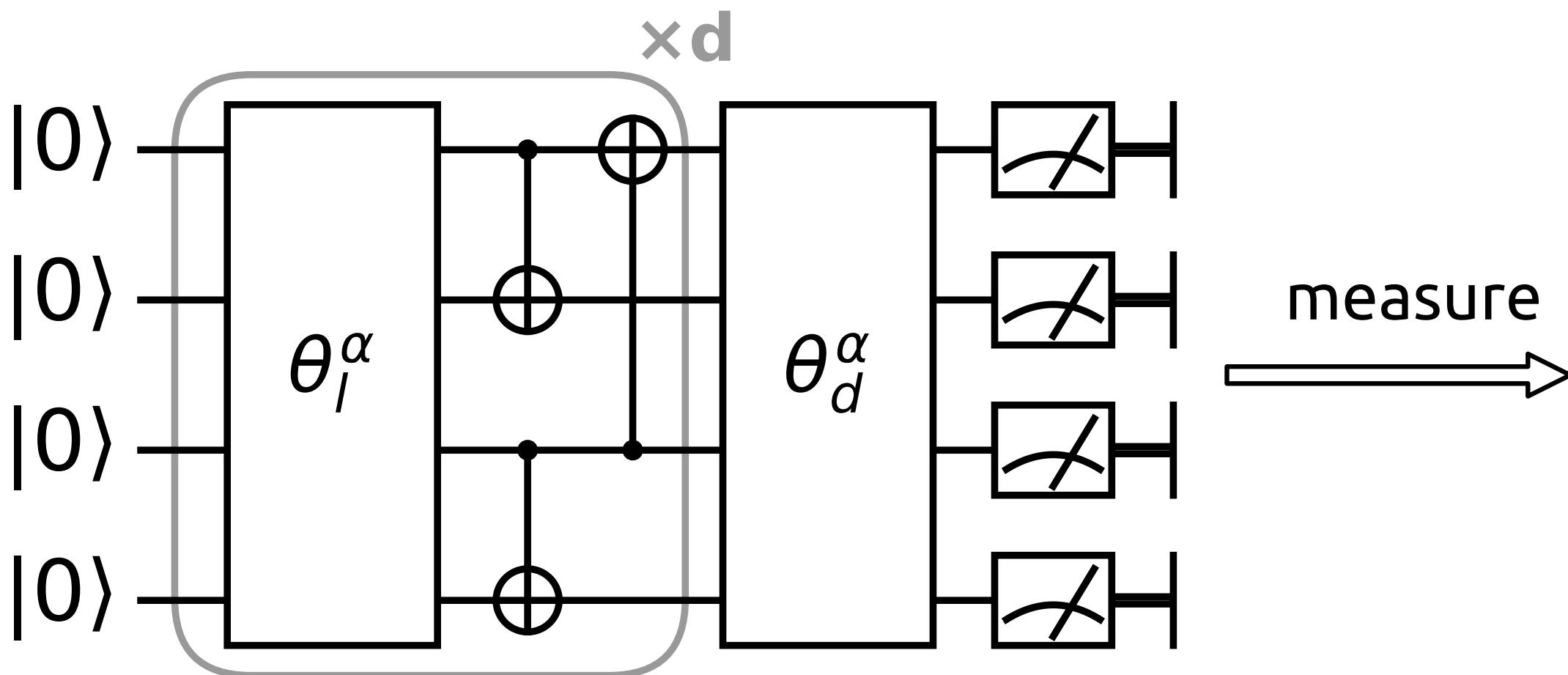
Hamiltonian engineering via differentiable programming

Differentiable Quantum Programming

With Liu, Zeng, Wu, Hu
1804.04168, 1808.03425

Short term:

What can we do with circuits of limited depth ?



Long term:

Are we really good at programming quantum computers ?

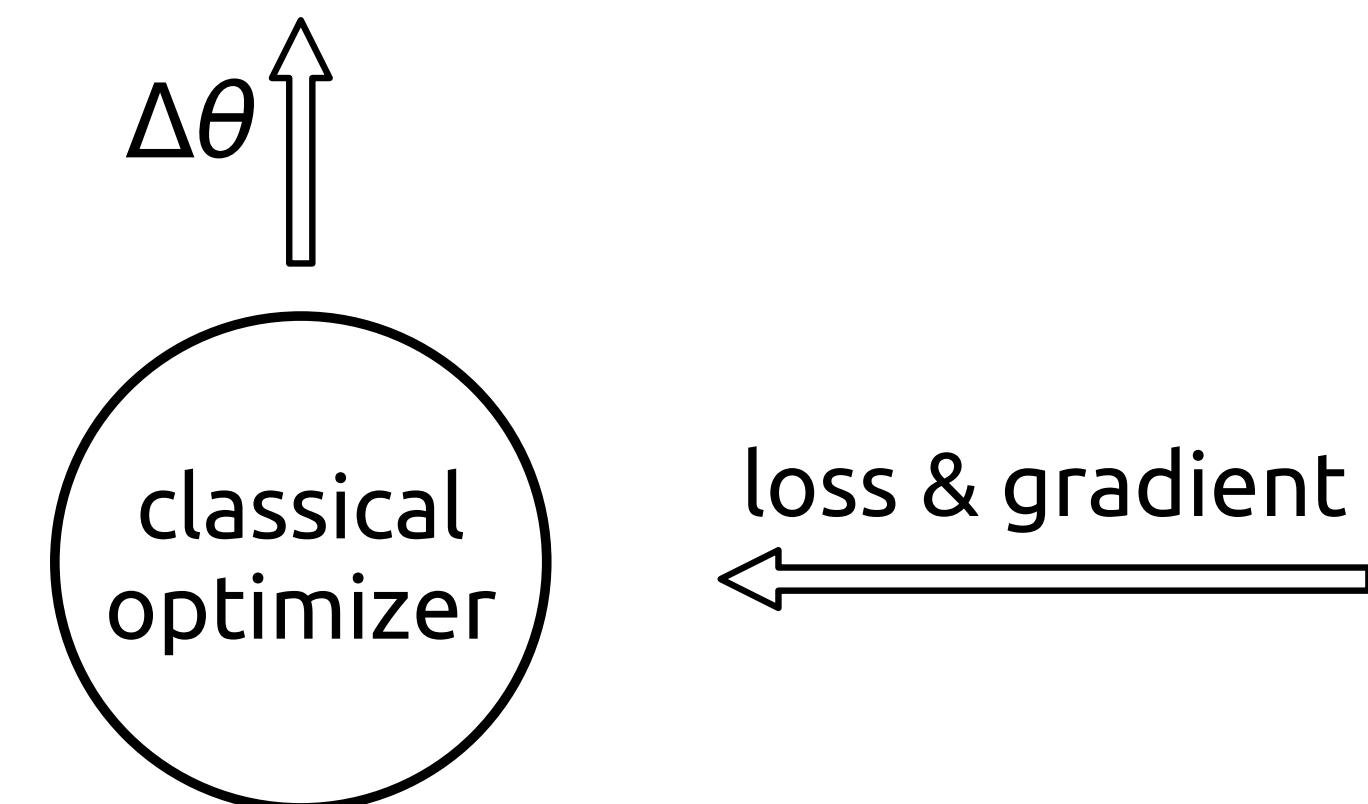
Quantum code



Andrey Karpathy ✅
@karpathy

Following

Gradient descent can write code better than you. I'm sorry.

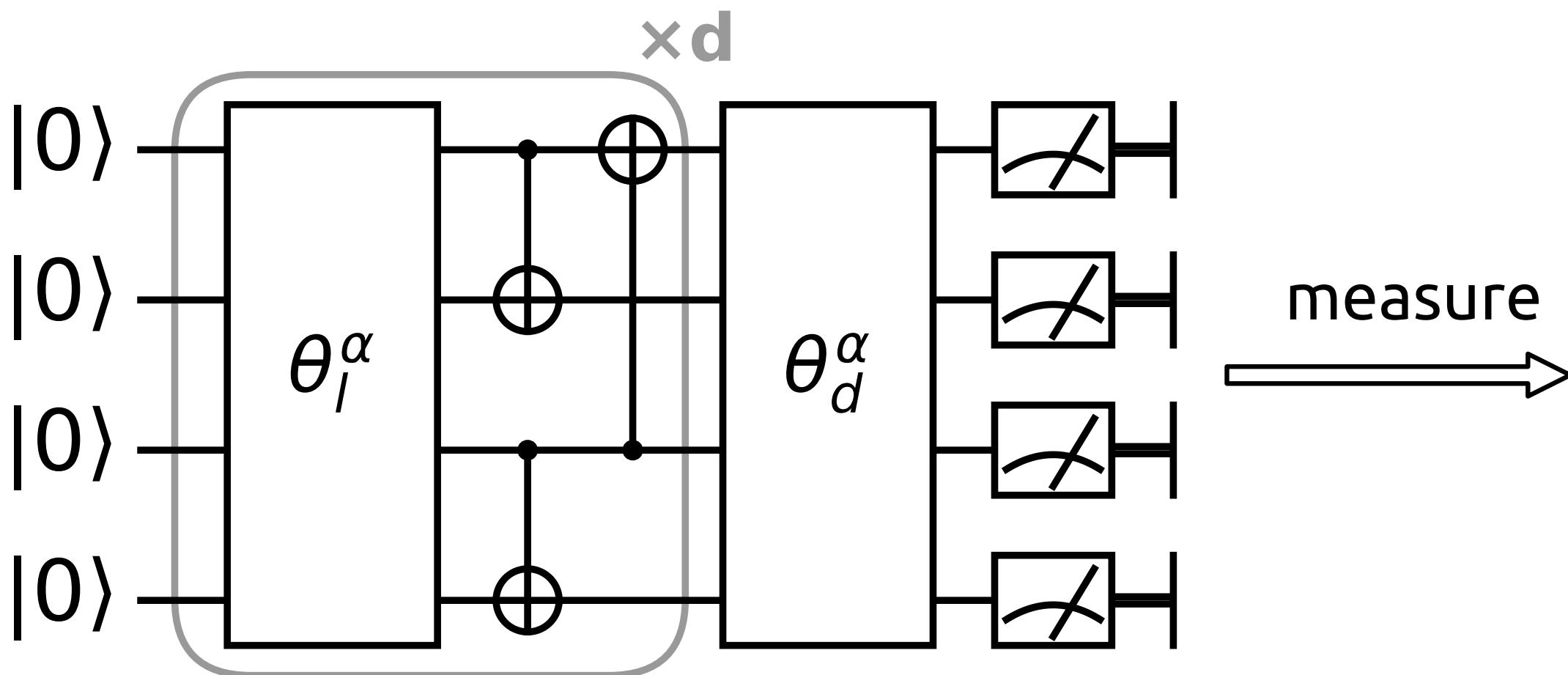


Differentiable Quantum Programming

With Liu, Zeng, Wu, Hu
1804.04168, 1808.03425

Short term:

What can we do with circuits of limited depth ?



Long term:

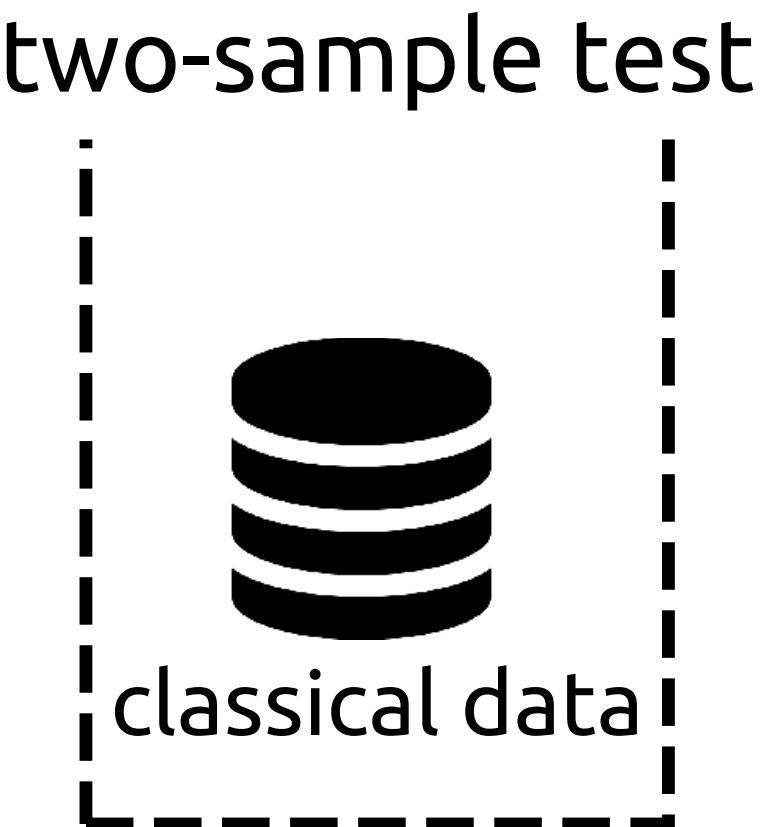
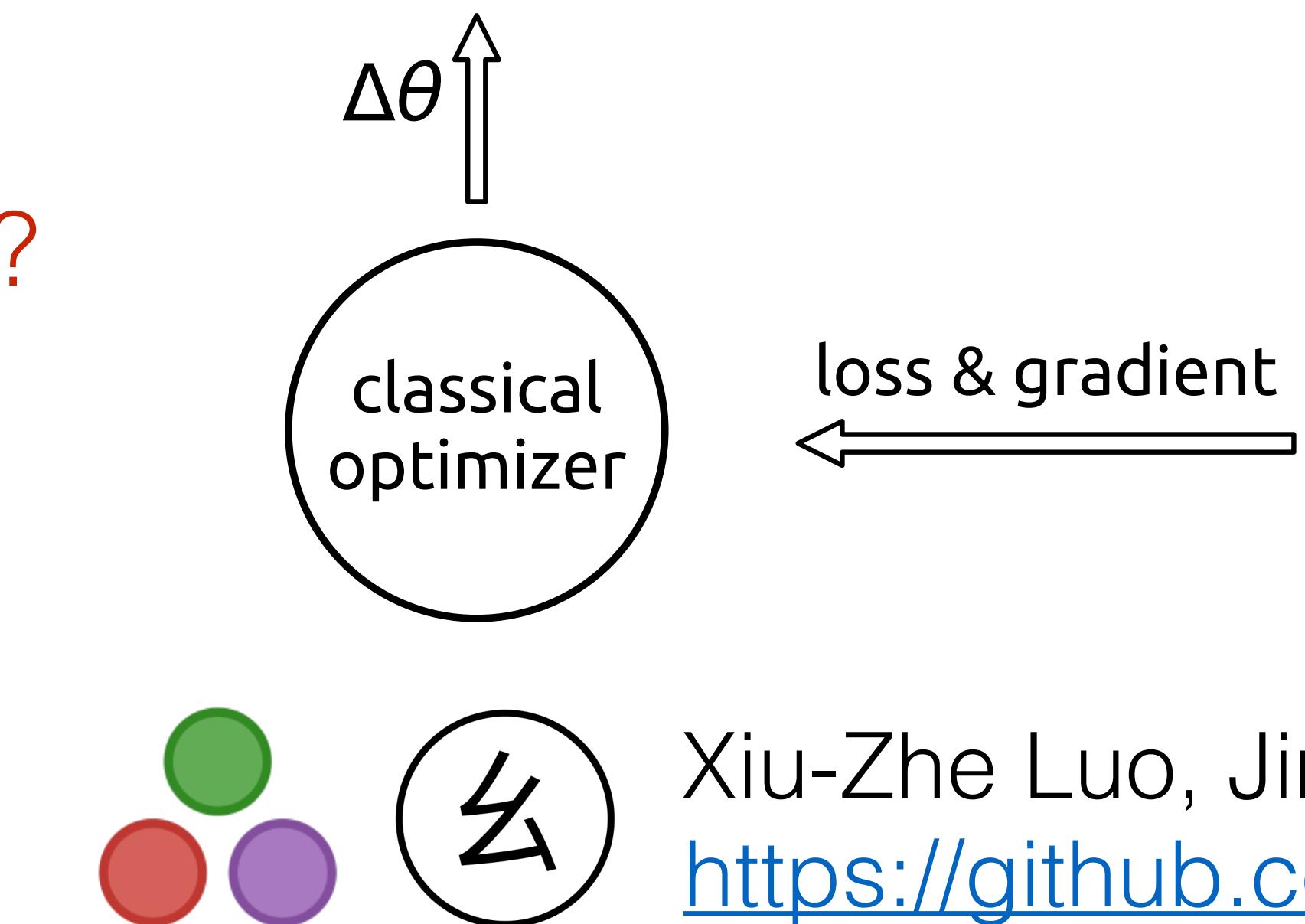
Are we really good at programming quantum computers ?

Quantum code



Andrej Karpathy ✅
@karpathy

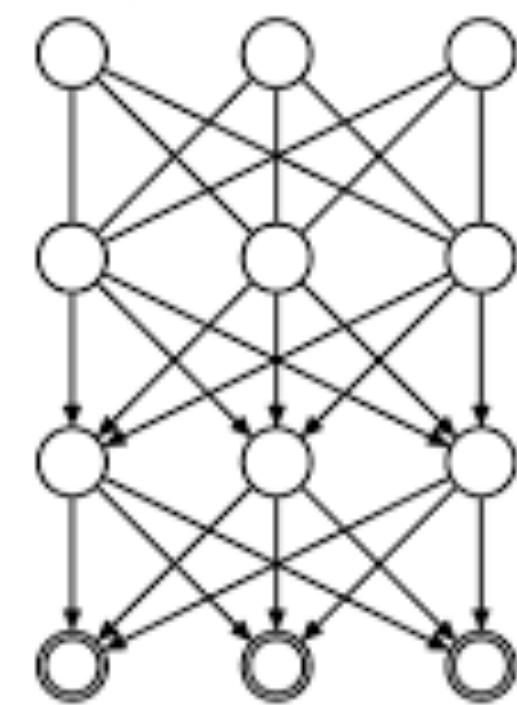
Gradient descent can write code better than you. I'm sorry.



Xiu-Zhe Luo, Jinguo Liu
<https://github.com/QuantumBFS/Yao.jl/>

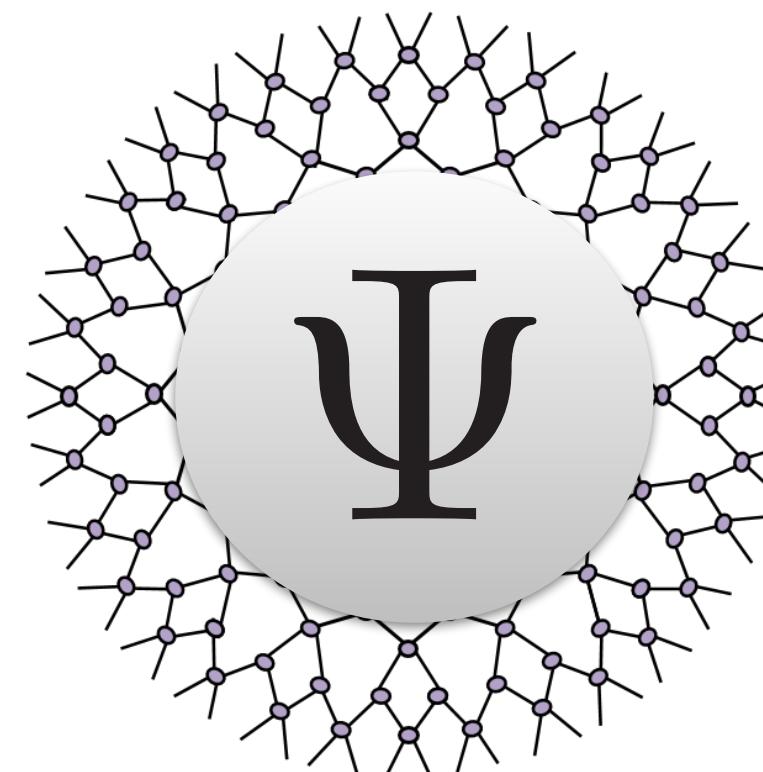
What is a deep neural network ?

Neural Net

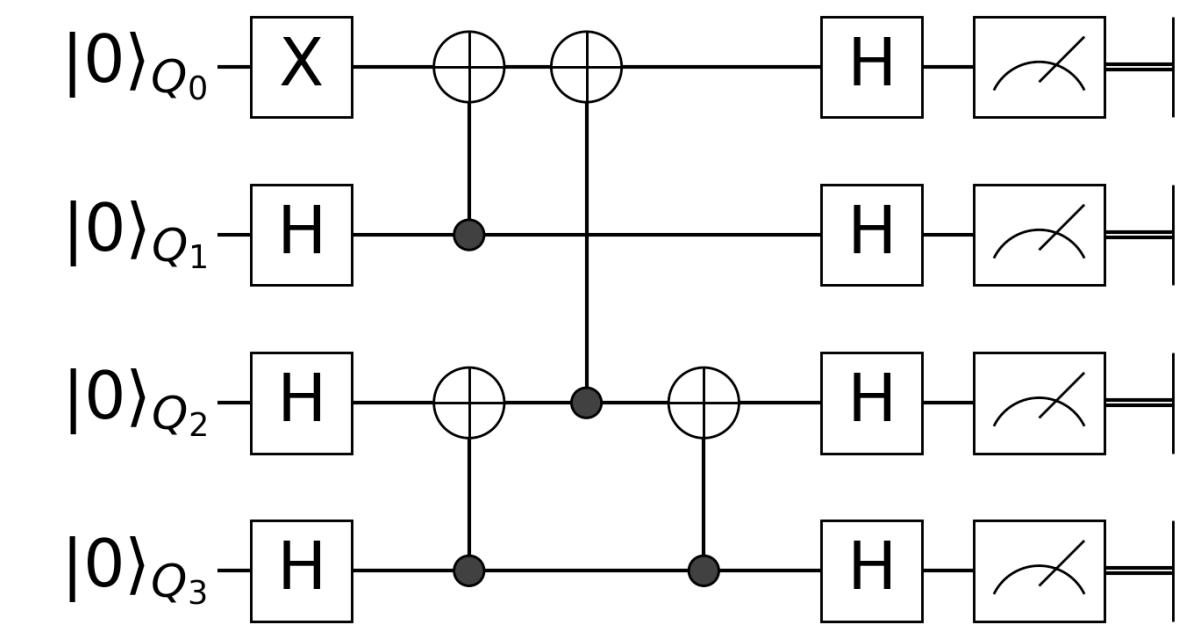


“三重境界”

Tensor Net



Quantum Circuit



1. Function Approximation
2. Probabilistic Transformation
3. Information Processing Device

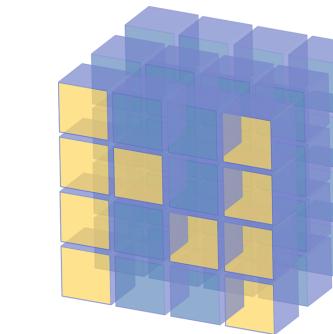
Hands on time!



<https://github.com/wangleiphy/dl4csrc>

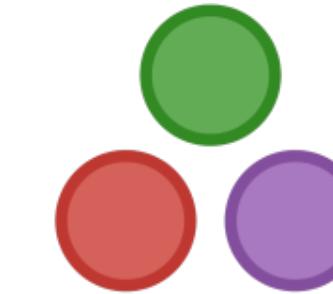
①

Back propagation from scratch



②

Differentiable Ising solver



③

Fun with normalizing flows



Thank You!

Jin-Guo Liu

Jinfeng Zeng

Xiu-Zhe Luo

Yufeng Wu

Pan Zhang

Dian Wu

Song Cheng

Shuo-Hui Li

Linfeng Zhang

Weinan E